Winter 12-15-2017

Image Reconstruction of the Speed of Sound and Initial Pressure Distributions in Ultrasound Computed Tomography and Photoacoustic Computed Tomography

Thomas Paul Matthews
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science
Department of Biomedical Engineering

Dissertation Examination Committee:
Mark A. Anastasio, Chair
Joseph P. Culver
Joseph A. O'Sullivan
Lihong V. Wang
Quing Zhu

Image Reconstruction of the Speed of Sound and Initial Pressure Distributions in Ultrasound Computed Tomography and Photoacoustic Computed Tomography

by

Thomas Paul Matthews

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2017
Saint Louis, Missouri

Table of Contents

List of Figures . . . . . **v**

List of Tables . . . . . **xi**

Acknowledgments . . . . . **xii**

Abstract . . . . . **xiv**

1 Introduction **1**
1.1 Overview and Motivation 1
1.2 Approach to Image Reconstruction 4
1.3 Outline of the Dissertation 5

2 Background **7**
2.1 Imaging physics 7
2.1.1 The acoustic wave equation 9
2.1.2 Continuous forward models 14
2.1.3 Discrete forward models 17
2.1.4 Numerical wave solvers 18
2.2 Image reconstruction 21
2.2.1 Estimation of the initial pressure in PACT 21
2.2.2 Estimation of the speed of sound in PACT 23
2.2.3 Estimation of the speed of sound in USCT 23

3 Reconstruction of the speed of sound in ultrasound tomography **26**
3.1 Overview 26
3.2 Dual averaging method 28
3.3 Description of computer-simulation studies 34
3.3.1 Methods 34
3.3.2 Image reconstruction 38
3.3.3 Bias-variance analysis 39
3.4 Results of computer-simulation studies 41
3.4.1 Images reconstructed by SGD 41
3.4.2 Images reconstructed by RDA 43
3.4.3 Images reconstructed by SGD and RDA 47

3.4.4 Bias-variance analysis . . . 51
3.5 Description of clinical studies . . . 53
3.6 Results of clinical studies . . . 56
3.7 Conclusions . . . 60

4 Joint reconstruction of the speed of sound and initial pressure distributions from PACT data alone . . . 62
4.1 Overview . . . 62
4.2 Parameterized speed of sound . . . 64
4.3 Joint reconstruction . . . 65
4.4 Description of computer-simulation studies . . . 68
4.4.1 Methods . . . 70
4.4.2 Fixed constant speed of sound . . . 73
4.4.3 Joint reconstruction . . . 73
4.4.4 Mass density variations . . . 75
4.4.5 Imperfect parameterization . . . 75
4.4.6 High-frequency pressure data . . . 76
4.5 Results of computer-simulation studies . . . 79
4.5.1 Fixed constant speed of sound . . . 79
4.5.2 Joint reconstruction . . . 80
4.5.3 Mass density variations . . . 82
4.5.4 Approximate segmentations . . . 85
4.5.5 Impact of frequency content . . . 86
4.6 Description of experimental studies . . . 87
4.6.1 Methods . . . 88
4.6.2 Fixed speed of sound . . . 89
4.6.3 Joint reconstruction . . . 90
4.7 Results of experimental studies . . . 91
4.8 Conclusions . . . 94

5 Joint reconstruction of the speed of sound and initial pressure distributions from combined PACT/USCT data . . . 96
5.1 Overview . . . 96
5.2 Joint reconstruction . . . 98
5.3 Description of computer-simulation studies . . . 100
5.3.1 Methods . . . 100
5.3.2 Reconstruction of the initial pressure from PACT data . . . 103
5.3.3 Reconstruction of the SOS from PACT data . . . 103
5.3.4 Reconstruction of the SOS from USCT data . . . 104
5.3.5 Reconstruction of initial pressure and SOS distributions using a sequential approach . . . 104
5.3.6 Joint reconstruction from combined PACT/USCT data . . . 105

5.4 Results of computer-simulation studies 107
5.4.1 Reconstruction of the initial pressure from PACT data 107
5.4.2 Reconstruction of the SOS from PACT data 108
5.4.3 Reconstruction of the SOS from USCT data 109
5.4.4 Reconstruction of initial pressure and SOS distributions using a sequential approach 110
5.4.5 Joint reconstruction from combined PACT/USCT data 111
5.5 Conclusions 121

6 Summary **123**

Appendix A Adjoint state method **125**

Appendix B Applications of the adjoint state method **129**

References **136**

List of Figures

2.1 Schematics of the data acquisition process for (a) PACT and (b) USCT. (© 2017 IOP Publishing) 8

3.1 A schematic of the measurement geometry. The measurement system consists of a circular ring array of ultrasonic transducers. These transducers are located in a larger rectangular simulation grid, over which the acoustic wave equation is solved. Within the ring array is a smaller rectangular region representing the reconstructed image. The estimated SOS distribution is calculated within the gray circular field-of-view within that region. (© 2017 IOP Publishing) 35

3.2 Sound speed distribution of (a) the numerical breast phantom and (b) the low-contrast two bar phantom employed in the bias-variance analysis, given in units of mm/μs. (© 2017 IEEE) 37

3.3 Plot of RMSE versus the number of iterations for images reconstructed by use of SGD with a constant step size of 0.1 for several regularization parameter values. (© 2017 IEEE) 41

3.4 Plot of RMSE versus (a) the number of iterations and (b) the number of wave equation solver runs for images reconstructed by use of SGD with a line search and with several constant step size values for a regularization parameter value of 5×10^{-4}. (© 2017 IEEE) 42

3.5 Images reconstructed by use of SGD using (a) a line search and (b) a constant step size of 0.1, shown after 500 iterations for a regularization parameter value of 5×10^{-4}. The insets in the upper right corners of each image correspond to a zoomed-in image of the inclusion located at about 7 o'clock. The larger images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs, while the insets are shown in a grayscale window of $[1.50, 1.53]$ mm/μs. (© 2017 IEEE) . . . 43

3.6 Images reconstructed by the unweighted RDA method with a fixed step size of 0.1 for regularization parameter values of (a) 1×10^{-5}, (b) 5×10^{-5}, (c) 1×10^{-4}, and (d) 5×10^{-4}, shown after 300 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE) 44

3.7 Plot of RMSE versus the number of iterations for (a) several regularization parameter values and a fixed step size of 0.1 and (b) several constant step size values and a fixed regularization parameter value of 1×10^{-4} for images reconstructed by use of the unweighted RDA method. (© 2017 IEEE) . . . 45

3.8 Images reconstructed by the weighted RDA method for regularization parameter values of (a) 1×10^{-5}, (b) 5×10^{-5}, (c) 1×10^{-4}, and (d) 5×10^{-4}, shown after 300 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE) 45
3.9 Images reconstructed by use of the unweighted RDA method with a fixed step size of 0.1 after (a) 20, (b) 50, (c) 100, and (d) 250 iterations. Images reconstructed by use of the weighted RDA method after (e) 20, (f) 50, (g) 100, and (h) 250 iterations. All results are shown for a regularization parameter value of 1×10^{-4} and in a grayscale window of $[1.47, 1.58]$ mm/μs. The RMSEs for each reconstructed image are displayed in the bottom left of each subfigure. (© 2017 IEEE) 46
3.10 (a) Profiles through y = -6.5 mm for reconstructed images obtained by use of the weighted RDA method and the unweighted RDA method with a fixed step size of 0.1, shown after (a) 20 iterations and (b) 250 iterations. (© 2017 IEEE) 47
3.11 Plot of RMSE vs. (a) the number of iterations and (b) the number of wave solver runs for the weighted and unweighted RDA methods. (© 2017 IEEE) 48
3.12 Images reconstructed by (Row 1) SGD with a constant step size of 0.1 and $\lambda = 5 \times 10^{-4}$; (Row 2) unweighted RDA with a fixed step size of 0.1 and $\lambda = 1 \times 10^{-4}$; (Row 3) SGD with a line search and $\lambda = 5 \times 10^{-4}$; and (Row 4) weighted RDA with $\lambda = 1 \times 10^{-4}$. The columns correspond to the images obtained, from left to right, after 20, 50, 100, and 250 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE) . . . 49
3.13 Profiles through y = -6.5 mm for images reconstructed by the use of SGD with a line search and weighted RDA. (© 2017 IEEE) 50
3.14 Plot of RMSE versus (a) the number of iterations and (b) the number of wave solver runs for SGD with a line search, SGD with a constant step size of 0.1, unweighted RDA with a step size of 0.1, and weighted RDA. (© 2017 IEEE) 51
3.15 Bias-variance curve for SGD with constant step size and the unweighted RDA method. The corresponding regularization parameter values are given by each point. (© 2017 IEEE) 52
3.16 Example reconstructed images from bias-variance analysis. (a) Image reconstructed by SGD with a regularization parameter value of 5×10^{-5}. (b) Image reconstructed by RDA with a regularization parameter value of 1×10^{-4}. The two images have approximately the same bias. Both images are shown in their full dynamic ranges. The SOS values are given in units of mm/μs. (© 2017 IEEE) 53
3.17 Initial estimate of the object reconstructed by use of an adjoint state method described in [6]. (© 2017 IEEE) 55

3.18 (Top row) Images reconstructed by use of SGD with a constant step size of 2.5×10^5 and regularization parameter values of (a) 1×10^{-10}, (b) 3×10^{-10}, (c) 1×10^{-9}, and (d) 3×10^{-9}. (Bottom row) Images reconstructed by use of the weighted RDA method with regularization parameter values of (e) 1×10^{-10}, (f) 3×10^{-10}, (g) 1×10^{-9}, and (h) 3×10^{-9}. Images are shown after 100 iterations and in a grayscale window of $[1.38, 1.60]$ mm/μs. (© 2017 IEEE) . 57
3.19 (Top row) Images reconstructed by use of SGD with a constant step size of 2.5×10^5 after (a) 5, (b) 20, (c) 50, and (d) 100 iterations with a regularization parameter value of 1×10^{-9}. (Bottom row) Images reconstructed by use of weighted RDA after (e) 5, (f) 20, (g) 50, and (h) 100 iterations with a regularization parameter value of 1×10^{-9}. All images are shown in a grayscale window of [1.38, 1.60] mm/μs. (© 2017 IEEE) 58
3.20 Images reconstructed by use of the weighted RDA method with a wavelet-based penalty and regularization parameter values of (a) 3×10^{-10}, (b) 1×10^{-9}, (c) 3×10^{-9}, and (d) 1×10^{-8}. Images are shown after 100 iterations and in a grayscale window of $[1.38, 1.60]$ mm/μs. (© 2017 IEEE) 59

4.1 (a) A schematic of the segmented tissue types within the mouse. The labels for each numbered tissue type are given in Table 4.1. Phantoms for (b) the normalized initial pressure distribution, given in arbitrary units, (c) the SOS distribution, given in units of mm/μs, and (d) the mass density distribution, given in units of mg/mm^3. In order to better visualize the soft tissue variations for the SOS distribution, the grayscale window for this phantom was set to [1.47, 1.7] mm/μs resulting in saturation of the bone, which has a SOS value of 3.198 mm/μs. 70
4.2 (a) The pressure amplitude and (b) frequency spectrum of the EIR employed in the main computer-simulation studies. 72
4.3 Estimated segmented regions for (a) a two-parameter SOS model and (b) a three-parameter SOS model. 76
4.4 Phantoms for (a) the initial pressure distribution, given in arbitrary units, and (b) the SOS distribution, given in units of mm/μs, for the high-frequency computer-simulation studies. 78
4.5 (a) The pressure amplitude and (b) frequency spectrum of the EIR employed in the high-frequency computer-simulation studies. 78
4.6 Reconstructed initial pressure distributions for $\lambda = 10^{-3}$ and fixed constant SOS values of (a) 1.480 mm/μs, (b) 1.490 mm/μs, (c) 1.500 mm/μs, and (d) 1.510 mm/μs. The images are shown in a grayscale window of [0.0, 1.1]. 79
4.7 Reconstructed initial pressure distributions for parameterized JR for $\lambda = 10^{-4}$ with (a) 1 parameter, (b) 2 parameters, (c) 3 parameters, (d) 7 parameters, and (e) 55869 parameters. The images are shown in a grayscale window of [0.0, 1.1]. 81

4.8 Reconstructed SOS distributions for parameterized JR for $\lambda = 10^{-4}$ with (a) 1 parameter, (b) 2 parameters, (c) 3 parameters, and (d) 7 parameters, and (e) 55869 parameters. The images are shown in a grayscale window of [1.47, 1.70] mm/μs. 83
4.9 Initial pressure distributions reconstructed by parameterized JR for $\lambda = 10^{-4}$ with 7 parameters from measured data that (a) ignored mass density variations and (b) included mass density variations. SOS distributions reconstructed by parameterized JR with 7 parameters from measured data that (c) ignored mass density variations and (d) included mass density variations. Mass density variations were not included in the reconstruction process. The initial pressure images are shown in a grayscale window of [0.0, 1.1]. The SOS images are shown in a grayscale window of [1.47, 1.70] mm/μs. 84
4.10 Reconstructed initial pressure distributions obtained by parameterized JR for $\lambda = 10^{-4}$ with (a) 2 parameters with perfect segmentation, (b) 2 parameters with imperfect segmentation, (c) 3 parameters with perfect segmentation, and (d) 3 parameters with imperfect segmentation. Reconstructed SOS distributions obtained by parameterized JR with (e) 2 parameters with perfect segmentation, (f) 2 parameters with imperfect segmentation, (g) 3 parameters with perfect segmentation, and (h) 3 parameters with imperfect segmentation. The initial pressure images are shown in a grayscale window of [0.0, 1.1]. The SOS images are shown in a grayscale window of [1.47, 1.70] mm/μs. 85
4.11 Normalized cost function values for (a) different tissue SOSs and (b) different background SOSs after applying a Hann low-pass filter to the measured data. Results are shown for several cutoff frequencies. 87
4.12 (a) The pressure amplitude and (b) frequency spectrum of the estimated EIR for the experimental system. 89
4.13 Estimated segmented regions of the mouse for a 2-parameter SOS model. . . 90
4.14 Reconstructed initial pressure distributions for several fixed constant SOS values of (a) 1.490 mm/μs, (b) 1.495 mm/μs, (c) 1.500 mm/μs, and (d) 1.505 mm/μs. Results are shown for $\lambda = 10^{-1}$ in a grayscale window of $[0, 8000]$. 92
4.15 Reconstructed initial pressure distributions for $\lambda = 10^{-1}$ assuming (a) the SOS distribution employed as the initial guess for the first stage of JR and (b) the SOS distribution obtained by parameterized JR for $Q = 2$. Results are shown in a grayscale window of $[0, 8000]$. 93
4.16 Zoomed-in region of the reconstructed initial pressure distributions for $\lambda = 10^{-1}$ assuming (a) a tuned constant SOS of 1.500 mm/μs and (b) the SOS distribution obtained by parameterized JR for $Q = 2$. The arrows point to structures that are in focus in the JR image, but not in the tuned constant SOS image. Results are shown in a grayscale window of $[0, 6000]$. 93

5.1 (a) The SOS distribution, given in units of mm/μs, and (b) the initial pressure distribution, given in arbitrary units, of the numerical breast phantom. (© 2017 IOP Publishing) 101
5.2 Initial pressure distributions reconstructed from noiseless data assuming (a) the true SOS distribution and (b) a constant SOS equal to the background SOS. (c) Profiles through the reconstructed images at $y = 14.25$ mm. Results are shown after 50 iterations. No regularization was employed. (© 2017 IOP Publishing) 108
5.3 Reconstructed SOS distribution from noiseless PACT data given (a) the true initial pressure distribution and (b) an initial pressure distribution shifted 1 mm in the x-direction from the true distribution. (© 2017 IOP Publishing) 109
5.4 Reconstructed SOS distributions from noisy USCT measurements for (a) 512 views and $\lambda_c = 3 \times 10^{-6}$, for (b) 32 views and $\lambda_c = 3 \times 10^{-7}$, for (c) 16 views and $\lambda_c = 3 \times 10^{-7}$, and for (d) 8 views and $\lambda_c = 1 \times 10^{-7}$. All reconstructed images are shown after 400 iterations and in a grayscale window of [1.467, 1.520] mm/μs. (© 2017 IOP Publishing) 110
5.5 (a) Reconstructed initial pressure distribution from noiseless data assuming the SOS distribution obtained from 8 noiseless USCT views. (b) Profile through the reconstructed image at $y = 14.25$ mm. The profile through the reconstructed initial pressure distribution assuming the true SOS distribution is shown as a reference. (© 2017 IOP Publishing) 111
5.6 SOS distributions reconstructed from combined noiseless PACT and USCT data for 512 USCT views for (a) $\beta = 10^{-1}$, (b) $\beta = 10^{3}$, and (c) $\beta = 10^{6}$. No regularization was employed. Results are shown after 2 outer iterations in a grayscale window of $[1.467, 1.520]$ mm/μs. (© 2017 IOP Publishing) 112
5.7 Convergence of (a) the cost function value for the initial pressure subproblem, (b) the cost function value for the SOS subproblem, (c) the initial pressure RMSE, and (d) the SOS RMSE for the case of combined noiseless PACT and USCT data for 8 USCT views. The parameters were $\lambda_c = 1 \times 10^{-3}$, $\lambda_p = 0$, and $\beta = 10^{4}$. For computational expediency, the value of $F_{SUS}(\mathbf{c})$ was evaluated only for a single realization of the encoding vector $\mathbf{w}$. (© 2017 IOP Publishing) 113
5.8 SOS distributions reconstructed from (a) noiseless USCT data and (b) noiseless PACT and USCT data for 16 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noisy data and 16 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 1 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 0$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^{4}$, $\lambda_c = 3 \times 10^{-4}$, and $\lambda_p = 0$. (© 2017 IOP Publishing) 115

5.9 SOS distributions reconstructed from (a) noiseless USCT data and (b) noiseless PACT and USCT data for 8 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noiseless data and 8 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 1 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 0$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^4$, $\lambda_c = 1 \times 10^{-3}$, and $\lambda_p = 0$. (© 2017 IOP Publishing) 117

5.10 SOS distributions reconstructed from (a) noisy USCT data and (b) noisy PACT and USCT data for 8 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noisy data and 8 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 3 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 3 \times 10^{-3}$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^3$, $\lambda_c = 1 \times 10^{-4}$, and $\lambda_p = 1 \times 10^{-2}$. (© 2017 IOP Publishing) 118

List of Tables

4.1 Parameter values for the initial pressure, SOS, and mass density distributions for each tissue type. The normalized initial pressure values are roughly based on the relative concentration of blood for each tissue type [33, 53, 68, 102]. . . 71
4.2 Different SOS parameterizations employed during image reconstruction. . . . 74
4.3 Values of different metrics that could be employed for selecting a constant SOS value. The optimal value according to each metric is shown in bold. . . 80
4.4 Estimated SOS values for different parameterized SOS models. For models with $Q < 7$, the estimated SOS values represent effective SOSs across several different tissue types. 82

5.1 RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noiseless data for 16 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing) 120
5.2 RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noiseless data for 8 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing) 120
5.3 RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noisy data for 8 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing) 120

Acknowledgments

I would like to start by thanking my Ph.D. advisor, Dr. Mark A. Anastasio, for all his support and guidance. I have benefited immensely from his teaching, his insights, his kindness, and his good humor. It has been my great pleasure to learn from him these past several years.

I am grateful to my other committee members, Dr. Joseph Culver, Dr. Joseph A. O'Sullivan, Dr. Lihong V. Wang, and Dr. Quing Zhu, for their thoughtful suggestions and support. I would also like to thank Dr. Kilian Weinberger for serving on my thesis committee prior to his move to Cornell University.

I would like to thank my collaborators, in particular Dr. Neb Duric and Dr. Lihong Wang. They have developed world-class imaging systems that enabled me to explore the ideas in this dissertation and validate them through experimental studies. In addition, I am grateful for our many discussions, which helped to guide this work.

I want to thank all the members of Dr. Anastasio's lab for many stimulating discussions and good memories. In particular, I would like to thank Kun Wang, who served as a mentor to me when I first joined Dr. Anastasio's lab.

I also owe a debt of gratitude to Karen Teasdale, who supported me during my transition to Dr. Anastasio's lab.

I would like to thank the Center for Biological Systems Engineering (CBSE) for their funding support and the Center for High Performance Computing (CHPC) for use of the computing cluster, which enabled many of the studies in this work.

Finally, I would like to thank my parents and my girlfriend Julia for their love and support.

Thomas Paul Matthews

Washington University in Saint Louis

December 2017

ABSTRACT OF THE DISSERTATION

Image Reconstruction of the Speed of Sound and Initial Pressure Distributions in Ultrasound Computed Tomography and Photoacoustic Computed Tomography

by

Thomas Paul Matthews

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2017

Professor Mark A. Anastasio, Chair

Ultrasound computed tomography (USCT) and photoacoustic computed tomography (PACT) are two emerging imaging modalities that have a wide range of potential applications from pre-clinical small animal imaging to cancer screening in human subjects. USCT is typically employed to measure acoustic contrasts, including the speed of sound (SOS) distribution, while PACT typically measures optical contrasts or some related quantity such as the initial pressure distribution. Their complementary contrasts and similar implementations make USCT and PACT a natural fit for a hybrid imaging system. Still, much work remains to realize this promise. First, USCT image reconstruction methods based on the acoustic wave equation, known as waveform inversion methods, are computationally burdensome, limiting their widespread use. Instead, image reconstruction methods based on geometric acoustics are often employed. These methods do not model higher-order diffraction effects and consequentially have poor resolution. In this dissertation, use of a novel stochastic optimization

method, which overcomes much of the computational burden of waveform inversion, is proposed. Second, most traditional PACT image reconstruction algorithms assume a constant SOS distribution. For many biological applications, this is a poor assumption that can result in reduced resolution, reduced contrast, and an increase in the number of imaging artifacts. More recent image reconstruction algorithms can compensate for a known heterogeneous SOS distribution; however, in practice, the SOS distribution is not known. Further, in general, the joint reconstruction (JR) of the SOS and initial pressure distributions from PACT measurements is unstable. Two methods are proposed to overcome this problem. In the first, a parameterized JR method is employed. Under this approach, the SOS distribution is assumed to have a known low-dimensional representation. By constraining the form of the SOS distribution, the JR problem can be made more stable. In the second method, few-view USCT measurements are added to the PACT data, and the initial pressure and SOS distributions are jointly estimated from the combined measurements. This approach effectively exploits acoustic information present in the PACT data, allowing both the initial pressure and SOS distributions to be more accurately reconstructed.

Chapter 1

Introduction

1.1 Overview and Motivation

Ultrasound computed tomography (USCT) is a promising imaging modality with numerous applications including breast cancer screening [17, 27, 39, 42, 64, 93, 96, 113]. USCT is ideally suited to breast imaging as it offers novel tissue contrasts that can help differentiate benign masses from tumors [39]. It has several potential advantages over conventional imaging methods, as it is radiation free, breast compression free, and relatively inexpensive. In addition, ultrasound imaging may offer some advantages over mammography for the detection of breast cancer in women with dense breasts [29, 56].

In USCT, a series of ultrasonic pulses are sent through the object-of-interest, and the resulting pressure wavefield is recorded by a collection of ultrasonic transducers. While USCT permits estimation of many possible acoustic contrasts, in this dissertation, the focus will be on estimation of the speed of sound (SOS) distribution.

There are several ways to estimate the SOS distribution in USCT. One set of methods, known as ray-based methods, are based on geometrical acoustics [5, 29, 43, 48, 80, 92, 96].

These methods fail to model higher-order diffraction effects and consequently can have poor resolution. Reconstruction algorithms based on the acoustic wave equation, known as waveform inversion methods, can produce high-resolution images, but are very computationally demanding [41, 60, 62, 85, 90, 113, 124].

Recently, a method called waveform inversion with source encoding (WISE) was developed, which eliminated much of this computational burden [57, 109, 124]. Under this approach, the SOS distribution is estimated by solving a stochastic optimization problem. While the stochastic formulation has many computational advantages, it also introduces additional challenges. In particular, regularization of the inverse problem may become less effective in the stochastic setting. Further, use of a line search to accelerate the image reconstruction method may result in less accurate reconstructed images. In Chapter 3, the regularized dual averaging method (RDA) is employed to overcome many of these challenges [72, 77, 116].

Photoacoustic computed tomography (PACT) is a hybrid imaging modality that combines optical excitation and acoustic detection [83, 110, 111]. It can provide both functional and structural information for pre-clinical and clinical applications, including small animal imaging [30, 31, 40, 58, 65, 114, 118] and human breast imaging [30, 32, 40, 58, 59, 117]. In PACT, the image contrast arises from the absorption of light by molecules within the object. Often, this light is in the form of a short laser pulse. The absorbed optical energy gives rise to pressure waves via the photoacoustic effect that are then recorded by ultrasonic transducers surrounding the object.

PACT has a number of advantages over other optical imaging techniques. In particular, since it allows use of diffuse light to excite the object and since acoustic scattering is much weaker in tissue than optical scattering, PACT has the ability to obtain high-resolution images at large penetration depths [83, 110, 111]. Additionally, PACT has advantages over

other non-optical functional imaging modalities: compared with MRI, it offers faster image acquisition at lower costs; and compared with PET and SPECT, it does not involve use of ionizing radiation [112].

Traditional image reconstruction methods for PACT assume that the medium is acoustically homogeneous. This assumption is often violated for biological and biomedical applications [45, 88, 120]. In particular, unaccounted for variations in the SOS distribution can give rise to artifacts in the reconstructed initial pressure distribution. More recently, several methods have been proposed to account for SOS variations, including time-reversal [44, 106, 121], iterative full-wave inversion [7, 46], and the Neumann-series or iterative time reversal approach [91, 101]. However, each of these methods requires some knowledge of the SOS distribution in order to accurately reconstruct the initial pressure distribution. In practice, the SOS distribution is typically unknown.

One solution is to try to jointly estimate the initial pressure and SOS distributions from the PACT data alone. This is unstable in general [47,100]. In Chapter 4, a method for mitigating this instability is introduced. This method relies on the use of *a priori* knowledge about the structure of the SOS distribution, namely that the SOS distribution has a low-dimensional parameterized representation. By constraining the SOS distribution according to this parameterization, accurate joint estimation of the initial pressure and SOS distributions can be performed, at least in some cases.

One alternative is try to estimate the SOS distribution by use of adjunct imaging data. As discussed earlier, USCT has previously been shown to allow accurate reconstruction of the SOS distribution for breast imaging [17, 27, 39, 42, 64, 93, 113]. Consequently, several groups have begun investigation of combined USCT/PACT imaging systems [31, 70, 115]. These systems offer the potential to obtain automatically co-registered images with both

optical and acoustic contrasts. These complementary contrasts could aid in a variety of medical imaging tasks, including breast cancer detection. In addition, the similar detection hardware should make integration of these modalities comparatively simple and inexpensive.

In cases where combined USCT/PACT systems have been employed, the SOS and initial pressure distributions were reconstructed in a sequential manner. First, the USCT data were employed to estimate the SOS distribution. Then, the estimated SOS distribution was employed when estimating the initial pressure distribution from the PACT data. As will be shown, this approach is not optimal. Since the photoacoustic waves propagate according to the acoustic properties of the medium, the measured PACT data contain information on the SOS distribution. This information is ignored under the sequential approach, which estimates the SOS distribution from USCT measurements alone. In Chapter 5, a novel approach that jointly estimates the initial pressure and SOS distributions from combined PACT and USCT measurements is proposed [73]. This synergistic approach effectively utilizes the acoustic information in the PACT data, allowing the SOS distribution to be accurately estimated from few-view USCT measurements. Further, it allows the initial pressure distribution to be more accurately estimated by providing an improved estimate of the SOS distribution compared to what could obtained from PACT measurements alone.

1.2 Approach to Image Reconstruction

The approach to image reconstruction adopted within this dissertation is based on two pillars: (1) the development of realistic physics-based models for describing imaging systems and (2) the use of optimization-based image reconstruction methods, which seek to minimize a chosen cost function. Typically, this cost function includes one or more terms that relate

to the distance between the measured data and the predicted data given the assumed model and one or more regularization terms that incorporate *a priori* knowledge about the object being imaged. While no mathematical model will perfectly describe an imaging system, the hope is that by improving the accuracy of the assumed model, more useful reconstructed medical images can be obtained. Similarly, the use of an optimization-based framework for image reconstruction is motivated by the desire to improve performance on medical imaging tasks through the use of constraints and regularization terms that are tailored to the particular problem at hand. The flexibility in choosing the imaging model, as well the opportunity to incorporate prior knowledge, may allow for improved performance over competing approaches.

1.3 Outline of the Dissertation

In Chapter 2, background information on USCT and PACT is provided. Models for the propagation of ultrasonic and photoacoustic waves, based on the acoustic wave equation, are described in their continuous and discrete forms. Existing image reconstruction methods based on these models are reviewed.

In Chapter 3, a novel stochastic optimization method for estimating the SOS distribution in USCT is introduced. This approach, based on the regularized dual averaging method, offers reduced image reconstruction times compared with conventional full waveform inversion methods while still producing high-resolution images. Much of this work was previously described in [72].

In Chapter 4, parameterized joint reconstruction of the initial pressure and SOS distributions from PACT data alone is investigated. Under this approach, a low-dimensional parameterized model for the SOS distribution is employed in order to stabilized the joint reconstruction problem.

In Chapter 5, a joint reconstruction method for estimating the initial pressure and SOS distributions from combined PACT and USCT data is developed. The impact of the number of USCT views on the accuracy of the estimated SOS distributions is investigated. Much of this work was previously described in [73].

The results of the dissertation are summarized in Chapter 6.

Chapter 2

Background

2.1 Imaging physics

In PACT, an object is illuminated with an optical source, typically a pulsed laser. The light is absorbed by molecules within the object. Some of this absorbed energy is converted into heat, leading to a small temperature rise and thermoelastic expansion. When the temporal width of the laser pulse is sufficiently short, this expansion can be viewed as giving rise to an initial pressure distribution [112]. This initial pressure distribution then propagates according to the acoustic wave equation and the resulting pressure wavefield is recorded by a collection of ultrasonic transducers surrounding the object [83, 112]. A schematic for this process is shown in Fig. 2.1a.

The initial pressure distribution is related to the optical properties of the medium by

$$p_0(\mathbf{r}) = \eta(\mathbf{r})\,\Gamma(\mathbf{r})\,A(\mathbf{r}), \tag{2.1}$$

where $p_0(\mathbf{r})$ is the initial pressure distribution, $A(\mathbf{r})$ is the absorbed optical energy density, $\eta(\mathbf{r})$ is the percentage of energy converted into heat, and $\Gamma(\mathbf{r})$ is the dimensionless

Grüneisen parameter [112]. The absorped optical energy density can be further divided into two components as

$$A(\mathbf{r}) = \mu_a(\mathbf{r})\,\Phi(\mathbf{r}), \tag{2.2}$$

where $\mu_a(\mathbf{r})$ is the optical absorption coefficient and $\Phi(\mathbf{r})$ is the fluence distribution produced by the optical source. While some image reconstruction methods for PACT seek to estimate $\mu_a(\mathbf{r})$, or even the concentrations of different molecules, in this dissertation, the focus is on the more modest task of estimating the initial pressure distribution. See [23] for a thorough review of the estimation of chromophore concentrations from the initial pressure distribution.

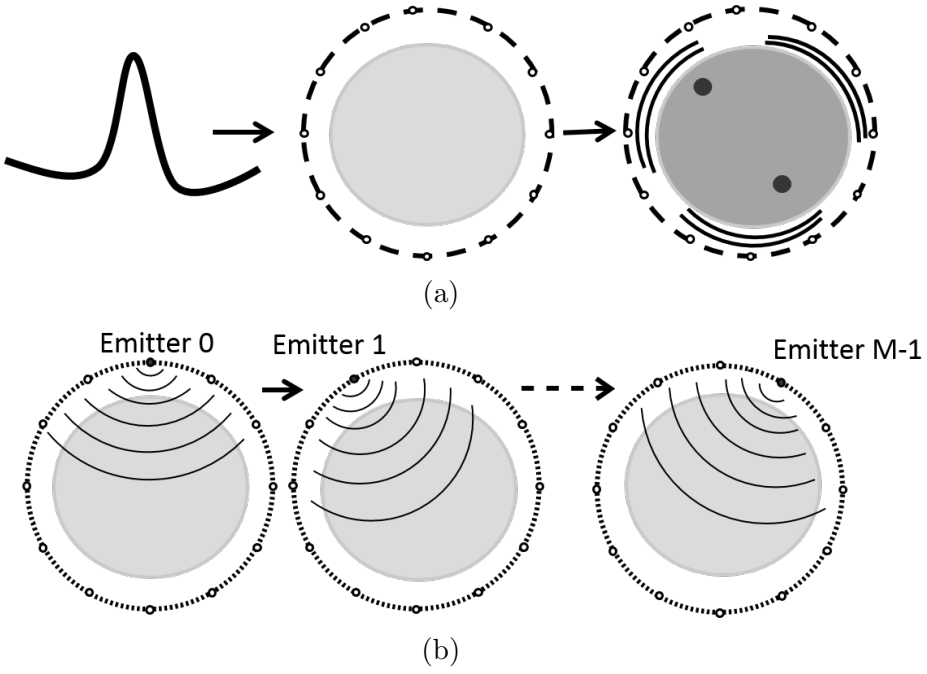


Figure 2.1: Schematics of the data acquisition process for (a) PACT and (b) USCT. (© 2017 IOP Publishing)

In USCT, an acoustic pulse is emitted from an ultrasonic transducer outside of the object. The generated acoustic wave then propagates through the object and the resulting wavefield is recorded by a collection of ultrasonic transducers. This process is then repeated with some

subset of the transducers serving as the emitter in turn (see Fig. 2.1b) [52, 61]. The goal is to estimate some acoustic property of the medium from this collection of measured pressure data. While many acoustic properties of the medium can potentially be estimated, in this dissertation, the focus is on the estimation of the SOS.

2.1.1 The acoustic wave equation

For both PACT and USCT, the propagation of acoustic waves is a key component of the image acquisition process. For this reason, the derivation of the acoustic wave equation, which describes the propagation of acoustic waves, is briefly reviewed (see [22, 54, 76, 86, 102, 112] for a more thorough treatment). As an acoustic wave propagates, it creates a perturbation in the material properties of the medium about some ambient values. As such, these quantities can be decomposed into two components: their ambient values, denoted with the subscript 0, and the perturbation about these ambient values due to the acoustic wave, denoted with the subscript a. For example, consider the corresponding equations for the pressure p, the particle velocity $\mathbf{u}$, and the mass density ρ:

$$p = p_0 + p_a \tag{2.3a}$$

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_a \tag{2.3b}$$

$$\rho = \rho_0 + \rho_a. \tag{2.3c}$$

Typically, one is most interested in the component that relates to the acoustic wave. Here, the focus is on fluid media (i.e. materials whose shear modulus and viscosity are equal to zero), which is often a sufficient analog for soft tissue.

The acoustic wave equation can be derived from three fundamental equations: (1) the continuity equation, which is a statement of the conservation of mass; (2) Euler's equation of motion for a fluid, which is an analog of Newton's Second Law for fluids; and (3) the pressure-density relation, which describes the thermodynamic relationship between the pressure and mass density for the material. Here, the focus will be on derivation of the linear acoustic wave equation, which assumes that the perturbation in the mass density due to the acoustic wave is small relative to the ambient mass density ($\rho_a \ll \rho_0$). First, consider the continuity equation, given by

$$\frac{\partial \rho\left(\mathbf{r},t\right)}{\partial t} = -\nabla \cdot \left(\rho\left(\mathbf{r},t\right)\mathbf{u}\left(\mathbf{r},t\right)\right). \tag{2.4}$$

This expression can be simplified by additionally assuming that the ambient mass density does not change ($\partial\rho_0/\partial t = 0$) and that there is no net flow of particles ($\mathbf{u}_0 = \mathbf{0}$):

$$\frac{\partial \rho_a\left(\mathbf{r},t\right)}{\partial t} = -\mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho\left(\mathbf{r},t\right) - \rho_0\left(\mathbf{r}\right)\nabla\cdot\mathbf{u}_a\left(\mathbf{r},t\right). \tag{2.5}$$

Expanding the expression for the mass density yields

$$\frac{\partial \rho_a\left(\mathbf{r},t\right)}{\partial t} = -\mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho_a\left(\mathbf{r},t\right) - \mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho_0\left(\mathbf{r},t\right) - \rho_0\left(\mathbf{r}\right)\nabla\cdot\mathbf{u}_a\left(\mathbf{r},t\right). \tag{2.6}$$

It can be shown that the first term on the right-hand side is negligible when he magnitude of the particle velocity is small relative to the SOS ($|\mathbf{u}| \ll c$) [22, 112]. In this case,

$$\frac{\partial \rho_a\left(\mathbf{r},t\right)}{\partial t} = -\mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho_0\left(\mathbf{r},t\right) - \rho_0\left(\mathbf{r}\right)\nabla\cdot\mathbf{u}_a\left(\mathbf{r},t\right). \tag{2.7}$$

Next, consider Euler's equation of motion for a fluid, given by

$$\rho\left(\mathbf{r},t\right)\frac{D\mathbf{u}\left(\mathbf{r},t\right)}{Dt}=-\nabla p\left(\mathbf{r},t\right), \tag{2.8}$$

where the total, or material, derivative D/Dt is defined as

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+\left(\mathbf{u}\cdot\nabla\right). \tag{2.9}$$

The total derivative describes changes in some property when the underlying medium may be moving [22,76]. For example, the mass density at some fixed position $\mathbf{r}$ could change both because the particles originally at that position have changed or because the particles that were originally at that position have been replaced by another collection of particles due to some flow of the medium. This second effect depends both on the flow of the material as well as the spatial variations in the underlying property. Assuming that the ambient pressure is homogeneous ($\nabla p_0=\mathbf{0}$) and as assumed previously $\mathbf{u}_0=\mathbf{0}$, $\rho_a\ll\rho_0$, and $|\mathbf{u}|\ll c$, Euler's equation can be simplified as

$$\rho_0\left(\mathbf{r}\right)\frac{\partial\mathbf{u}_a\left(\mathbf{r},t\right)}{\partial t}=-\nabla p_a\left(\mathbf{r},t\right). \tag{2.10}$$

When the underlying medium is in thermodynamic equilibrium (aside from the small perturbation caused by the acoustic wave itself), the pressure can be expressed solely as a function of the mass density and the entropy s of the medium [76,86]:

$$p=p\left(\rho,s\right). \tag{2.11}$$

When the entropy is constant in the sense that $Ds/Dt = 0$, the pressure and mass density are related by

$$\frac{Dp\left(\mathbf{r},t\right)}{Dt} = \left(\frac{\partial p}{\partial \rho}\right)_s \frac{D\rho\left(\mathbf{r},t\right)}{Dt}, \tag{2.12}$$

where the subscript s indicates the case of constant entropy. Under these conditions, a thermodynamic definition of the SOS can be given as

$$c_0^2 \equiv \left(\frac{\partial p}{\partial \rho}\right)_s. \tag{2.13}$$

The pressure-density relation can be further simplified by assuming the ambient pressure is time-independent ($\partial p_0/\partial t = 0$) and by use of the previously stated assumptions,

$$\frac{\partial p_a\left(\mathbf{r},t\right)}{\partial t} + \mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla p_a\left(\mathbf{r},t\right) = c_0^2\left(\mathbf{r}\right)\left(\frac{\partial \rho_a\left(\mathbf{r},t\right)}{\partial t} + \mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho\left(\mathbf{r},t\right)\right). \tag{2.14}$$

Dropping the terms that are non-linear in the acoustic field variables yields,

$$\frac{\partial p_a\left(\mathbf{r},t\right)}{\partial t} \approx c_0^2\left(\mathbf{r}\right)\left(\frac{\partial \rho_a\left(\mathbf{r},t\right)}{\partial t} + \mathbf{u}_a\left(\mathbf{r},t\right)\cdot\nabla\rho_0\left(\mathbf{r},t\right)\right). \tag{2.15}$$

Due to this approximation, the resulting acoustic wave equation is often referred to as the linear acoustic wave equation. When the mass density is homogeneous, the pressure-density relation has the simpler, more familiar, form:

$$p_a\left(\mathbf{r},t\right) \approx c_0^2\left(\mathbf{r}\right)\rho_a\left(\mathbf{r},t\right). \tag{2.16}$$

Solving for the time-derivative of the acoustic mass density and substituting the resulting expression into Eqn. 2.7 gives

$$\frac{1}{c_0^2(\mathbf{r})}\frac{\partial p_a(\mathbf{r},t)}{\partial t} = -\rho_0(\mathbf{r})\,\nabla\cdot\mathbf{u}_a(\mathbf{r},t)\,. \tag{2.17}$$

Note that the two terms involving the gradient of the ambient mass density cancel. Combining this with Eqn. 2.10 gives a pair of differential equations that describe the propagation of acoustic waves,

$$\rho_0(\mathbf{r})\frac{\partial \mathbf{u}_a(\mathbf{r},t)}{\partial t} = -\nabla p_a(\mathbf{r},t) \tag{2.18a}$$

$$\frac{1}{\rho_0(\mathbf{r})\,c_0^2(\mathbf{r})}\frac{\partial p_a(\mathbf{r},t)}{\partial t} = -\nabla\cdot\mathbf{u}_a(\mathbf{r},t)\,. \tag{2.18b}$$

Due to the assumptions on the entropy and viscosity, these equations neglect acoustic attenuation. Further, in cases where the acoustic properties of the medium are discontinuous, it may be necessary to solve the above equations as a series of boundary value problems as described in [50]. For the SOS and mass density distributions considered in this dissertation, this difficulty is avoided.

While Eqn. 2.18 is given in terms of the particle velocity and the acoustic pressure, for the image reconstruction tasks considered in this dissertation, it is sufficient to model the pressure, which is the quantity most directly recorded by the ultrasonic transducers. In this case, the two equations can be combined to yield a single differential equation, which in some

cases may be simpler to solve:

$$\frac{1}{\rho_0(\mathbf{r})\, c_0^2(\mathbf{r})} \frac{\partial^2 p_a(\mathbf{r},t)}{\partial t^2} = \nabla \cdot \left(\frac{1}{\rho_0(\mathbf{r})} \nabla p_a(\mathbf{r},t) \right) \tag{2.19a}$$

$$p_a(\mathbf{r},0) = 0 \tag{2.19b}$$

$$\frac{\partial p_a}{\partial t}(\mathbf{r},0) = 0. \tag{2.19c}$$

To distinguish between the two forms, this wave equation will be referred to as the second-order wave equation as it involves a second-order temporal derivative, while the pair of coupled first-order differential equations will be referred to as the first-order wave equation. When the mass density is spatially homogeneous, Eqn. 2.19 can be further simplified to

$$\frac{1}{c_0^2(\mathbf{r})} \frac{\partial^2 p_a(\mathbf{r},t)}{\partial t^2} = \nabla^2 p_a(\mathbf{r},t). \tag{2.20}$$

In the rest of this dissertation, for cleanness of notation, the subscripts on the SOS, mass density, acoustic pressure, and particle velocity will be suppressed.

2.1.2 Continuous forward models

By use of the linear acoustic wave equation, models for the data acquisition process, or forward models, for both PACT and USCT can be developed. To start, the models will be described in terms of the first-order acoustic wave equation.

The propagation of the photoacoustic waves through the medium is described by

$$\rho(\mathbf{r})\frac{\partial \mathbf{u}(\mathbf{r},t)}{\partial t}+\nabla p(\mathbf{r},t)=\mathbf{0} \tag{2.21a}$$

$$\frac{1}{\rho(\mathbf{r})\,c(\mathbf{r})^{2}}\frac{\partial p(\mathbf{r},t)}{\partial t}+\nabla\cdot\mathbf{u}(\mathbf{r},t)=0 \tag{2.21b}$$

$$p(\mathbf{r},0)=p_0(\mathbf{r}) \tag{2.21c}$$

$$\mathbf{u}(\mathbf{r},0)=\mathbf{0}, \tag{2.21d}$$

where $p(\mathbf{r},t)$ is the acoustic pressure, $\mathbf{u}(\mathbf{r},t)$ is the particle velocity, $c(\mathbf{r})$ is the SOS, and $p_0(\mathbf{r})$ is the initial pressure distribution.

Similarly, a forward model describing the data acquisition process for USCT can be developed as

$$\rho(\mathbf{r})\frac{\partial \mathbf{u}_m(\mathbf{r},t)}{\partial t}+\nabla p_m(\mathbf{r},t)=\mathbf{0} \tag{2.22a}$$

$$\frac{1}{\rho(\mathbf{r})\,c(\mathbf{r})^{2}}\frac{\partial p_m(\mathbf{r},t)}{\partial t}+\nabla\cdot\mathbf{u}_m(\mathbf{r},t)=4\pi\int_0^t dt' s_m(\mathbf{r},t') \tag{2.22b}$$

$$p_m(\mathbf{r},0)=0 \tag{2.22c}$$

$$\mathbf{u}_m(\mathbf{r},0)=\mathbf{0}, \tag{2.22d}$$

where the subscript m denotes the m-th view and $s_m(\mathbf{r},t)$ is the excitation pulse emitted by the ultrasonic transducer acting as the emitter for the m-th view. These differential equations differ from those for PACT only in terms of the initial conditions and the acoustic source term. In the case of homogeneous mass density, the equivalent forward model based

on the second-order acoustic wave equation is given by

$$\nabla^2 p_m(\mathbf{r},t) - \frac{1}{c(\mathbf{r})^2}\frac{\partial^2 p_m(\mathbf{r},t)}{\partial t^2} = -4\pi s_m(\mathbf{r},t) \tag{2.23a}$$

$$p_m(\mathbf{r},0) = 0 \tag{2.23b}$$

$$\frac{\partial p_m}{\partial t}(\mathbf{r},0) = 0. \tag{2.23c}$$

The data measured by the ultrasonic transducers acting as receivers can be obtained from the pressure over the whole domain as

$$g_i(t) = \mathcal{M}_i p(\mathbf{r},t), \tag{2.24}$$

where $g_i(t)$ is the data recorded by the i-th receiver and $\mathcal{M}_i$ is an operator that gives the data recorded by the i-th transducer from the data over the whole domain. In this dissertation, the operator $\mathcal{M}_i$ is chosen to have the form

$$\mathcal{M}_i \equiv h_e(t) *_t \Lambda_i, \tag{2.25}$$

where Λ_i is the restriction of the pressure over the whole domain to the location of the i-th receiver and $h_e(t)$ is the electro-acoustic impulse response (EIR), which describes the frequency dependence of the ultrasonic transducer. Here, for simplicity, all of the ultrasonic transducers are assumed to have the same EIR. In some cases, the frequency dependence of transducers may be ignored, which is equivalent to choosing $h_e(t) = \delta(t)$. The spatial impulse responses of the transducers are not considered in this dissertation.

2.1.3 Discrete forward models

While digital imaging systems are most naturally described by continuous-to-discrete imaging models [10], for the purposes of iterative image reconstruction, it is often necessary to develop a discrete-to-discrete (D-D) model. Such D-D models are also needed for certain numerical methods for solving the acoustic wave equation.

The D-D forward model for PACT can be denoted as

$$\mathbf{g}_{PA} = \mathbf{M}\mathbf{H}_{PA}\left(\mathbf{c}\right)\mathbf{p}_0, \tag{2.26}$$

where $\mathbf{p}_0 \in \mathbb{R}^N$ is a discrete representation of the initial pressure distribution $p_0\left(\mathbf{r}\right)$, $\mathbf{c} \in \mathbb{R}^N$ is a discrete representation of the SOS $c\left(\mathbf{r}\right)$, $\mathbf{H}_{PA}\left(\mathbf{c}\right) \in \mathbb{R}^{NL \times N}$ is an operator that gives a discrete approximation of the solution of the wave equation specified in Equation 2.21. Here, a pixel-basis is employed to represent the SOS and the initial pressure distributions. As will be shown, this representation allows for an efficient numerical solution to the acoustic wave equation. Thus, $N \equiv N_x N_y$ is the total number of pixels, and N_x and N_y are the number of pixels along the x- and y-dimensions, respectively. Additionally, $\mathbf{M} \in \mathbb{R}^{N_{rec}L \times NL}$ is an operator that gives the pressure recorded by the transducers from the pressure over the whole domain (i.e. it is a concatenation of the discrete approximations of $\mathcal{M}_i$ for all transducers), $\mathbf{g}_{PA} \in \mathbb{R}^{N_{rec}L}$ is the measured data for all receivers, and N_{rec} is the number of transducers acting as receivers.

The D-D forward model for USCT can be denoted as

$$\mathbf{g}_m = \mathbf{M}\mathbf{H}_{US}\left(\mathbf{c}\right)\mathbf{s}_m, \tag{2.27}$$

where $\mathbf{s}_m \in \mathbb{R}^{NL}$ is a discrete approximation of the excitation pulse for the m-th view, $\mathbf{H}_{US}(\mathbf{c}) \in \mathbb{R}^{NL \times NL}$ is an operator that gives a discrete approximation to the solution to the acoustic wave equation given in either Eqn. 2.22 or Eqn. 2.23 depending on the chosen numerical method, and $\mathbf{g}_m \in \mathbb{R}^{N_{rec}L}$ is the measured data for the m-th view.

2.1.4 Numerical wave solvers

In order for the proposed forward models to be computationally practical, efficient methods for solving the acoustic wave equations given in Equations 2.18 and 2.19 are needed. The k-space pseudospectral methods fulfill this requirement [71, 103]. These time-domain methods calculate the acoustic pressure (and possibly the particle velocity) at a series of time points in a sequential manner starting from some initial conditions.

A key computational advantage of the k-space pseudospectral methods is their ability to accurately simulate the acoustic pressure with very few spatial samples per wavelength compared with finite-difference-based methods [35, 71, 103]. This is achieved by calculating the spatial derivatives in the acoustic wave equation using the following property of the Fourier transform

$$\mathcal{F}\left\{\frac{\partial f}{\partial x}\right\} = jk_x \mathcal{F}\{f\}, \tag{2.28}$$

where $\mathcal{F}$ is the Fourier transform, $f(x)$ is some function, j is the imaginary number, and k_x is the Fourier variable corresponding to x [35, 104].

Another distinguishing feature of the k-space pseudospectral method, compared with other pseudospectral methods, is the use of a non-standard finite-difference scheme for the temporal derivatives. This scheme is chosen such that, prior to discretization, the obtained solution to the difference-based acoustic wave equation is exact for the case of an acoustically homogeneous medium. As an example, the update step for the second-order acoustic wave equation is given below [71]:

$$\frac{Q(\mathbf{k}, t+\Delta t) - 2Q(\mathbf{k}, t) + Q(\mathbf{k}, t-\Delta t)}{\Delta t^2 sinc^2\left(kc_{ref}\Delta t/2\right)} = -c_{ref}^2 k^2 P(\mathbf{k}, t) + 4\pi c_{ref}^2 S(\mathbf{k}, t), \tag{2.29}$$

where Δt is the time step, $\mathbf{k}$ is the Fourier variable corresponding to $\mathbf{r}$, c_{ref} is some constant reference SOS, and

$$P\left(\mathbf{k}, t\right) \equiv \mathcal{F}_{2D}\left\{p(\mathbf{r}, t)\right\} \tag{2.30a}$$

$$S\left(\mathbf{k}, t\right) \equiv \mathcal{F}_{2D}\left\{s(\mathbf{r}, t)\right\} \tag{2.30b}$$

$$Q\left(\mathbf{k}, t\right) \equiv \mathcal{F}_{2D}\left\{\left(c_{ref}^2 / c(\mathbf{r})^2\right) p(\mathbf{r}, t)\right\}, \tag{2.30c}$$

where $\mathcal{F}_{2D}\left\{\cdot\right\}$ is the 2D spatial Fourier transform. The sinc term in the denominator of the left-hand side provides a correction to the approximation of the second-order temporal derivative that helps reduce the numerical error for large time steps. This correction factor approaches one as the time step approaches zero. More detailed information on the k-space pseudospectral method can be found in [71, 103].

While the first-order and second-order acoustic wave equations, given in Equations 2.18 and 2.19, are equivalent in a continuous setting, there are some important practical differences between the numerical methods employed to solve each equation. In particular, both of those equations describe free-space propagation, but in a discrete setting, the simulation

grid is necessarily finite. Without care, erroneous reflections or wrap-around effects can occur near the boundaries of the simulation grid. For the implementation of the first-order k-space pseudospectral method, a perfectly matched layer (PML) is introduced in order to absorb, without reflection, the pressure waves as they approach the boundaries of the simulation grid. The PML employed in this dissertation is based on a split-field formulation originally proposed for modeling the electromagnetic wave equation [13]. As part of this approach, the pressure field must be split into x- and y-components. Together with the x- and y-components of the particle velocity, this yields four different field variables that must be updated at each time step compared with one for the second-order k-space formulation. However, in its original form, the second-order k-space pseudospectral method does not include a PML [71]. Thus, the simulation grid must be sufficiently large to avoid any errors in the measured pressure due to wrap-around effects. For the purposes of image reconstruction, only errors in the simulated acoustic field that are recorded by the transducers are significant. Therefore, the relative trade-off between these two approaches will depend on the size of the transducer array relative to the pixel size as well as the record time for each of the transducers.

The choices of the temporal and spatial sampling rates have a large impact on the computational effort required to solve the acoustic wave equation. Choosing these rates to be too fine will result in wasted computational effort, but choosing them to be too coarse can result in numerical instability and large errors in the simulated pressure. The level of error that can be tolerated will depend on the specific task for which the numerical wave solver is being employed. In the case of image reconstruction, small numerical errors may not be the dominate source of artifacts in the reconstructed images. Other sources of error, such as model error, may play a more dominate role. In this dissertation, the following rules of thumb are roughly employed. The spatial sampling interval is chosen to be between $\lambda_{min}/3$ and

$\lambda_{min}/2.5$, where λ_{min} is the smallest wavelength expected within the simulation grid. The temporal sampling rate is then chosen so that the Courant-Friedrichs-Lewy (CFL) number, defined as $c_0\Delta t/\Delta x$ where Δx is the pixel size, is between 0.3 and 0.5 [71, 104].

2.2 Image reconstruction

While there are many possible image reconstruction methods for both PACT and USCT, this dissertation will focus on use of optimization-based image reconstruction methods [36, 37]. These methods provide a flexible framework for image reconstruction that can accurately model the underlying physics of the imaging system.

2.2.1 Estimation of the initial pressure in PACT

An estimate of the initial pressure distribution can be obtained by solving the penalized least-squares optimization problem

$$\widehat{\mathbf{p}_0} = \underset{\mathbf{p}_0 \geq 0}{\arg\min}\, F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right) + \lambda R\left(\mathbf{p}_0\right), \tag{2.31}$$

where $F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right)$ is a data fidelity term, $R\left(\mathbf{p}_0\right)$ is a regularization term, and λ is a regularization parameter, which controls the relative weight of the two terms. Here, $F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right)$ is given by

$$F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right) = \frac{1}{2}\|\underline{\mathbf{g}_{PA}} - \mathbf{M}\mathbf{H}_{PA}\left(\mathbf{c}\right)\mathbf{p}_0\|_2^2, \tag{2.32}$$

where $\underline{\mathbf{g}_{PA}}$ is measured data, which may be inconsistent with the chosen imaging model due to noise or model error. For this problem, the data fidelity term is differentiable and convex with respect to $\mathbf{p}_0$. When the regularization function is chosen to be convex, Eqn. 2.31 represents a convex optimization problem since the constraint $\mathbf{p}_0 \geq \mathbf{0}$ defines a convex set. In this dissertation, $R(\mathbf{p}_0)$ is chosen to be the total variation (TV) semi-norm, given by

$$R_{TV}(\mathbf{p}_0) = \|\nabla \mathbf{p}_0\|_1, \tag{2.33}$$

which is convex, but non-smooth.

While there are many possible optimization methods for solving Eqn. 2.31, in this dissertation, the Fast Iterative Shrinkage / Thresholding Algorithm (FISTA) is employed. This method has a number of advantages over alternative first-order optimization methods (i.e. methods that require knowledge of the gradient, but not higher-order derivatives). First, it permits the use of non-smooth regularization terms, such as the TV semi-norm. Second, it belongs to a family of algorithms that for weakly convex optimization problems achieves the optimal asymptotic convergence rate. Namely, for a weakly convex function $F(\mathbf{x})$, FISTA obtains the convergence rate

$$F(\mathbf{x}_k) - \min_{\mathbf{x}} F(\mathbf{x}) \sim O\left(1/k^2\right), \tag{2.34}$$

where k is the iteration number and $\mathbf{x}_k$ is the estimate of the sought-after quantity for the k-th iteration. The problem given by Eqn. 2.31 represents a weakly convex optimization problem whenever $\mathbf{MH}_{PA}$ is not full rank and $\mathcal{R}(\mathbf{p}_0)$ is convex.

2.2.2 Estimation of the speed of sound in PACT

While not typically done, it is possible to estimate the SOS distribution from PACT data alone. In this dissertation, to investigate the sensitivity of the estimated SOS distribution to the assumed initial pressure distribution, the following optimization problem is considered

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right) + \lambda \mathcal{R}\left(\mathbf{c}\right). \tag{2.35}$$

Errors in the estimated SOS distribution induced by an inaccurate assumed initial pressure distribution may provide some insight into the stability of the JR problem (see [47] for a detailed investigation). In order to provide flexibility in the type of optimization method employed, a smoothed version of the TV penalty is utilized,

$$\mathcal{R}_{STV}\left(\mathbf{c}\right) = \sum_{i,j} \sqrt{\left(\left[\mathbf{c}\right]_{i,j} - \left[\mathbf{c}\right]_{i-1,j}\right)^2 + \left(\left[\mathbf{c}\right]_{i,j} - \left[\mathbf{c}\right]_{i,j-1}\right)^2 + \epsilon}, \tag{2.36}$$

where $\left[\mathbf{c}\right]_{i,j}$ denotes the $(iN_x + j)$-th element of $\mathbf{c}$ and $\epsilon > 0$ is a small smoothing parameter used to prevent division by zero when calculating the gradient. Here, we choose $\epsilon = 10^{-12}$.

2.2.3 Estimation of the speed of sound in USCT

For USCT, the SOS distribution can be estimated by full-waveform inversion as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} F_{US}\left(\mathbf{c}\right) + \lambda \mathcal{R}\left(\mathbf{c}\right), \tag{2.37}$$

where

$$F_{US}(\mathbf{c}) \equiv \frac{1}{2}\sum_{m=0}^{M-1} \|\underline{\mathbf{g}_m} - \mathbf{M}\mathbf{H}_{US}(\mathbf{c})\,\mathbf{s}_m\|_2^2 \tag{2.38}$$

and $\underline{\mathbf{g}_m} \in \mathbb{R}^{N_{rec}L}$ is the measured data for m-th view and M is the total number of views. This approach can produce more accurate and higher-resolution estimated SOS distributions than ray-based methods [109], but can be computationally expensive as evaluating the cost function requires the wave equation to be solved M times. Recently, a source encoding technique has been employed to efficiently estimate the SOS distribution in USCT [57, 109, 124]. Using this technique, the SOS distribution can be estimated as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \mathbf{E}_{\mathbf{w}} \left\{ \frac{1}{2}\|\underline{\mathbf{g}_w} - \mathbf{M}\mathbf{H}_{US}(\mathbf{c})\,\mathbf{s}_w\|_2^2 \right\} + \lambda\mathcal{R}(\mathbf{c}), \tag{2.39}$$

where $\mathbf{w}$ is a random encoding vector, chosen to according to a Rademacher distribution [57, 107], $\mathbf{E}_{\mathbf{w}}$ is the expectation with respect to $\mathbf{w}$, and

$$\underline{\mathbf{g}_w} = \sum_{m=0}^{M-1} [\mathbf{w}]_m\, \underline{\mathbf{g}_m} \quad \text{and} \quad \mathbf{s}_w = \sum_{m=0}^{M-1} [\mathbf{w}]_m\, \mathbf{s}_m \tag{2.40}$$

are the encoded measurement data and the encoded source term, respectively. Under this approach, rather than simulating the pressure wavefield for each emitter separately, the pressure is simulated for the case where all emitters are fired simultaneously. For each emitter, the excitation pulse is scaled by the corresponding element of $\mathbf{w}$. Since the acoustic wave equation is linear with respect to the source term, the measured data can be similarly encoded. When $\mathbf{w}$ has zero mean and an identity covariance matrix, the optimization problems given in Eqns. 2.37 and 2.39 are equivalent [107]. For brevity, this new data fidelity term

will be denoted as

$$F_{SUS}\left(\mathbf{c}\right) \equiv \mathbf{E}_{\mathbf{w}}\left\{\frac{1}{2}\|\underline{\mathbf{g}_{w}} - \mathbf{M}\mathbf{H}_{US}\left(\mathbf{c}\right)\mathbf{s}_{w}\|_{2}^{2}\right\}. \tag{2.41}$$

This data fidelity term is non-convex with respect to the SOS distribution. As a result, the initial guess for $\mathbf{c}$ can have a large impact on the estimated SOS distribution.

Previously, the optimization problem given in Eqn. 2.39 was solved by use of stochastic gradient descent (SGD) [109]. This approach involves choosing a single realization of $\mathbf{w}$ for each iteration of the SGD method, greatly reducing the number of times that the acoustic wave equation must be solved.

The task of estimating the value of Eqn. 2.41 from limited samples of $\mathbf{w}$ can be related to randomized trace estimation [107]. While many choices for the distribution of $\mathbf{w}$ will produce unbiased estimates of $F_{US}\left(\mathbf{c}\right)$, different distributions may produce estimates with different variances. In this dissertation, the Rademacher distribution is chosen as it produces low variance estimates when compared with some alternative distributions [8]; however, in practice, it was observed that several different choices for the distribution of $\mathbf{w}$ yielded similar accuracies [107].

Chapter 3

Reconstruction of the speed of sound in ultrasound tomography

3.1 Overview

As discussed in Chapter 2, waveform inversion methods for USCT image reconstruction can produce high-resolution SOS images, but have high computational cost. An approach that combines waveform inversion with source encoding and alleviates much of the computational burden was previously proposed [57, 109, 124]. This approach reframes the image reconstruction problem as a stochastic optimization problem where the data fidelity term is the expectation of a random quantity.

In [109], this optimization problem was solved by use of SGD. Under this approach, the stochastic data fidelity term and the deterministic regularization term are treated jointly as part of a single cost function. This approach has several limitations. First, it fails to exploit the structure of the objective function [116]. In other words, SGD treats the cost function as a black box, ignoring potentially useful information about the nature of the cost

function. For example, in Eqn. 2.39, the cost function consists of two terms: a stochastic but differentiable data fidelity term and a deterministic regularization term. In SGD, this knowledge is ignored, and the gradients of the stochastic and deterministic terms are lumped together. Second, it assumes that all terms in the cost function are differentiable. This is not true of many sparsity-promoting regularization functions, such as the ℓ_1-norm and the total variation (TV) semi-norm. In some cases, the non-smooth regularization term can be approximated by a smoothed differentiable version through the introduction of a small smoothing parameter [109]. While this approach can be effective in some cases, the strategy could be difficult to apply to certain non-smooth regularization functions. Third, it fails to exploit information from previous iterations. For SGD, at each iteration, only the gradient corresponding to a single realization of the encoding vector is considered when determining the search direction. When combined with a line search for choosing the step size, this can lead to overfitting [95]. In this case, the line search method will choose a large step size that effectively minimizes the cost function evaluated for a single realization of the encoding vector, but which increases, or less effectively minimizes, the cost function evaluated for a large number of realizations. This problem can be overcome by the use of a fixed step size, at the expense of slowing the convergence rate.

Here, we propose use of a structured optimization method, known as the regularized dual averaging method (RDA), that considers the two terms in the cost function separately [77,116]. This approach can mitigate the impact of the stochastic data fidelity on the deterministic regularization term and result in more effective regularization that offers superior trade-offs between image resolution and noise variance by exploiting the structure of the cost function. It also provides the opportunity to employ non-smooth penalties in the waveform inversion cost function, which can be designed to exploit certain sparseness properties of the object [9, 15, 99].

3.2 Dual averaging method

The dual averaging method is a primal-dual optimization method originally developed by Nesterov [77] and later extended by Xiao to include regularization [116]. It can be employed to solve stochastic optimization problems of the same form as given in Eqn. 2.39. Here, we review the RDA method and detail its application to waveform inversion. For clarity, we do not attempt to describe the most general form of the RDA method, but merely one that has proven effective for waveform inversion. This application differs from the class of problems originally considered by Xiao and Nesterov in that the data fidelity term of our cost function is non-convex [77, 116].

For context, SGD is briefly reviewed [98]. When solving Eqn. 2.39, the update step for the $(k+1)$-th iteration of SGD can be written as

$$\mathbf{c}^{(k+1)} = \arg\min_{\mathbf{c}} \left\{ \left\langle \mathbf{G}_{US}^{(k)}, \mathbf{c} \right\rangle + \frac{1}{2\alpha_k} \|\mathbf{c} - \mathbf{c}^{(k)}\|_2^2 + \lambda \mathcal{R}\left(\mathbf{c}\right) \right\}, \tag{3.1}$$

or equivalently,

$$\mathbf{c}^{(k+1)} = \mathbf{c}^{(k)} - \alpha_k \left(\mathbf{G}_{US}^{(k)} + \lambda \nabla_{\mathbf{c}} \mathcal{R}\left(\mathbf{c}^{(k)}\right) \right), \tag{3.2}$$

where $\mathbf{c}^{(k)}$ is the estimate of the SOS distribution at the k-th iteration, α_k is the step size, $\langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner product, and

$$f_{US}\left(\mathbf{c}, \mathbf{w}\right) \equiv \frac{1}{2} \|\underline{\mathbf{g}_w} - \mathbf{M}\mathbf{H}_{US}\left(\mathbf{c}\right) \mathbf{s}_w\|_2^2 \tag{3.3}$$

$$\mathbf{G}_{US}^{(k)} \equiv \nabla_{\mathbf{c}} f_{US}\left(\mathbf{c}^{(k)}, \mathbf{w}^{(k)}\right), \tag{3.4}$$

where $\nabla_{\mathbf{c}}$ is the gradient with respect to $\mathbf{c}$. A more detailed summary of SGD is given in Algorithm 1.

Algorithm 1 Stochastic gradient descent (SGD)

Input: $\mathbf{c}^{(0)}$, λ
Output: $\hat{\mathbf{c}}$

$k \leftarrow 0$ {k is the algorithm iteration number.}
while stopping criterion is not satisfied **do**
    Draw $\mathbf{w}^{(k)}$ according to chosen distribution.
    Calculate $\mathbf{G} \leftarrow \nabla_{\mathbf{c}} f_{US}\left(\mathbf{c}^{(k)}, \mathbf{w}^{(k)}\right) + \lambda \nabla_{\mathbf{c}} \mathcal{R}\left(\mathbf{c}^{(k)}\right)$
    Choose step size α_k
    $\mathbf{c}^{(k+1)} \leftarrow \mathbf{c}^{(k)} - \alpha_k \mathbf{G}$
    $k \leftarrow k + 1$
end while
$\hat{\mathbf{c}} \leftarrow \mathbf{c}^{(k)}$

For the RDA method described in Algorithm 2, the update step for the $(k+1)$-th iteration is given by

$$\mathbf{c}^{(k+1)} = \arg\min_{\mathbf{c}} \left\{ \left\langle \overline{\mathbf{G}}_{US}^{(k)}, \mathbf{c} \right\rangle + \frac{1}{2\mu_k} \|\mathbf{c} - \mathbf{c}^{(0)}\|_2^2 + \lambda \mathcal{R}\left(\mathbf{c}\right) \right\}, \tag{3.5}$$

where $\overline{\mathbf{G}}_{US}^{(k)}$ is the average gradient of the data fidelity term over all past iterations, and $\mu_k > 0$ is a scalar. This update step is different from the update step for SGD in two key ways. First, the average gradient across iterations is employed instead of the gradient for a single iteration. By averaging across iterations, some of the stochastic noise associated with the data fidelity term can be mitigated without increasing the per-iteration computational cost of the algorithm. Second, the proximal term, $\frac{1}{2}\|\mathbf{c} - \mathbf{c}^{(0)}\|_2^2$, is independent of the iteration number. Together these two differences allow the RDA method to utilize non-local information when determining the estimate of the object for the next iteration.

In the case of simple averaging, the average gradient of the data fidelity term is given by

$$\overline{\mathbf{G}}_{US}^{(k)} = \frac{1}{k+1} \sum_{i=0}^{k} \nabla_{\mathbf{c}} f\left(\mathbf{c}^{(i)}, \mathbf{w}^{(i)}\right). \tag{3.6}$$

A weighted average of gradient estimates can also be considered as [77]

$$\overline{\mathbf{G}}_{US}^{(k)} = \frac{1}{\sum_{i=0}^{k} \alpha_i} \sum_{i=0}^{k} \alpha_i \nabla_{\mathbf{c}} f\left(\mathbf{c}^{(i)}, \mathbf{w}^{(i)}\right), \tag{3.7}$$

where $\{\alpha_i\}_{i=0}^{k}$ are weights for each of the gradient estimates. Here, the weights are chosen using a line search via a method analogous to existing line search procedures in conventional gradient descent. However, the proposed line search procedure for the weights does differ in some important ways from the conventional approach. Since the search direction for the RDA method is given by the average gradient, adjusting the current weight will determine the contribution of the most recent gradient estimate to the search direction. As a result, the search direction changes as the weight is adjusted. In conventional line search methods, the search direction is fixed and only the magnitude of the update is adjusted. Further, as the search direction for the RDA method incorporates information from multiple iterations corresponding to different realizations of the encoding vector, it is less prone to overfitting than SGD with a line search.

When the regularization function $\mathcal{R}$ is convex, the dual averaging update step can be written in terms of the proximity operator of $\mathcal{R}$ as

$$\mathbf{c}^{(k+1)} = \text{prox}_{\lambda\mu_k \mathcal{R}}\left(\mathbf{c}^{(0)} - \mu_k \overline{\mathbf{G}}_{US}^{(k)}\right), \tag{3.8}$$

where the proximity operator is defined as [84]

$$\text{prox}_{\lambda R}(\mathbf{x}) \equiv \min_{\mathbf{y}} \left\{ R(\mathbf{y}) + \frac{1}{2\lambda}\|\mathbf{x}-\mathbf{y}\|^2 \right\} \tag{3.9}$$

for $\lambda > 0$. Proximity operators are a common way to handle non-smooth terms in an optimization problem. Many existing works give expressions and procedures for evaluating the proximal operator for different common regularization functions [11, 19, 21, 119]. Unless otherwise noted, the regularization function is chosen to be the TV semi-norm as the TV semi-norm has been shown to be effective at mitigating noise while preserving sharp edges [18]. While the proximity operator of the TV semi-norm has no closed form expression, it can be efficiently computed using the fast gradient projection method described in [11, 19]. Under this approach, the computational cost of applying the proximity operator is much less than that of computing the gradient; so the computational cost of the RDA method is approximately the same as SGD on a per-iteration basis.

From Eqn. 3.8, it can be seen that the update step for the dual averaging method can be divided into two parts. First, a reference value, or initial guess, for the sought-after object is updated based on a weighted sum of all past gradient estimates of the data fidelity term. Second, regularization is incorporated by use of the associated proximity operator of the regularization term. In this way, the stochastic data fidelity term is handled separately from the deterministic regularization term. Additionally, averaging the gradients of the data fidelity term obtained over several iterations may help minimize the impact of the variance of the gradient due to the random encoding vector.

In the RDA method, the weights of the individual gradient realizations are separate from the distance to step along the search direction. This later responsibility is handled by the

Algorithm 2 Regularized dual averaging (RDA) method

Input: $\mathbf{c}^{(0)}$, λ
Output: $\hat{\mathbf{c}}$

1: $k \leftarrow 0$ {k is the algorithm iteration number.}
2: $A_{-1} \leftarrow 0$
3: **while** stopping criterion is not satisfied **do**
4: Draw $\mathbf{w}^{(k)}$ according to chosen distribution.
5: Calculate $\mathbf{G}_{US}^{(k)} \leftarrow \nabla_{\mathbf{c}} f_{US}\left(\mathbf{c}^{(k)}, \mathbf{w}^{(k)}\right)$
6: Choose weight $\alpha_k > 0$ {Unweighted case: $\alpha_k = 1$}
7: $A_k \leftarrow A_{k-1} + \alpha_k$
8: $\overline{\mathbf{G}}_{US}^{(k)} \leftarrow \left(1 - \frac{\alpha_k}{A_k}\right) \overline{\mathbf{G}}_{US}^{(k-1)} + \frac{\alpha_k}{A_k} \mathbf{G}_{US}^{(k)}$ {Compute weighted average of gradient.}
9: Choose μ_k {For example, $\mu_k = \gamma A_k$, where $\gamma > 0$ is a constant.}
10: $\mathbf{c}^{(k+1)} \leftarrow \mathbf{c}^{(0)} - \mu_k \overline{\mathbf{G}}_{US}^{(k)}$
11: $\mathbf{c}^{(k+1)} \leftarrow \text{prox}_{\lambda \mu_k \mathcal{R}}\left(\mathbf{c}^{(k+1)}\right)$
12: $k \leftarrow k + 1$
13: **end while**
14: $\hat{\mathbf{c}} \leftarrow \mathbf{c}^{(k)}$

sequence $\{\mu_k\}$. For simplicity, we choose $\mu_k = \gamma A_k$, where $A_k = \sum_{i=0}^{k} \alpha_i$ and $\gamma > 0$ is a constant. In the this case, line 10 in Algorithm 2 reduces to

$$\mathbf{c}^{(k+1)} \leftarrow \mathbf{c}^{(0)} - \gamma \sum_{i=0}^{k} \alpha_i \mathbf{G}_{US}^{(i)}. \tag{3.10}$$

Under this formulation, γ plays the role of the step size, and the unweighted RDA method corresponds to the case where $\alpha_k = 1$. As with step sizes in other optimization methods, the constant γ should be chosen to be sufficiently small to insure convergence. For example, when the Lipschitz constant of the gradient of the data fidelity term is known, γ could be chosen to be the inverse of the Lipschitz constant in the unweighted case or the inverse of the product of the Lipschitz constant and the maximum allowable weight of the gradient α_{max} in the weighted case.

The line search procedure for the weighted RDA method is described by Alg. 3. The goal of this procedure is to find weights that improve the convergence rate of the algorithm while minimizing the computational cost needed to select those weights. The goal is not necessarily to choose weights that most minimize the cost function at each iteration as this might be a less efficient use of computational resources than updating the estimate of the object more regularly with less precisely chosen weights. Each weight value considered for a given iteration requires the cost function to be evaluated one additional time. Since the data fidelity term is evaluated for only one realization of the encoding vector, this requires only one additional wave solver run. This computational cost is the same as for the line search procedure employed for SGD.

Algorithm 3 Line search for RDA method

Input: $\mathbf{c}^{(0)}$, A_{k-1}, $\mathbf{w}^{(k)}$, $\mathbf{G}_{US}^{(k)}$, $\overline{\mathbf{G}}_{US}^{(k-1)}$, $f_{US}\left(\mathbf{c}^{(k)}, \mathbf{w}^{(k)}\right)$, λ, α_{max}
Output: α_k {Weight for k-th iteration.}

$\tilde{\alpha} \leftarrow \alpha_{max}$ {α_{max} is the initial guess for the weight.}
$found \leftarrow$ **false**
while not $found$ **do**
    $\tilde{A} \leftarrow A_{k-1} + \tilde{\alpha}$
    $\widetilde{\mathbf{G}} \leftarrow \left(1 - \frac{\tilde{\alpha}}{\tilde{A}}\right) \overline{\mathbf{G}}_{US}^{(k-1)} + \frac{\tilde{\alpha}}{\tilde{A}} \mathbf{G}_{US}^{(k)}$
    $\tilde{\mu} \leftarrow \gamma \tilde{A}$ {Should be consistent with Alg. 2.}
    $\tilde{\mathbf{c}} \leftarrow \text{prox}_{\lambda\tilde{\mu}\mathcal{R}}\left(\mathbf{c}_0 - \tilde{\mu}\widetilde{\mathbf{G}}\right)$
    if $f_{US}\left(\tilde{\mathbf{c}}, \mathbf{w}^{(k)}\right) + \lambda\mathcal{R}\left(\tilde{\mathbf{c}}\right) < f_{US}\left(\mathbf{c}^{(k)}, \mathbf{w}^{(k)}\right) + \lambda\mathcal{R}\left(\mathbf{c}^{(k)}\right)$ **then**
        $found \leftarrow$ **true**
    else
        $\tilde{\alpha} \leftarrow \tilde{\alpha}/2$
    end if
end while
$\alpha_k \leftarrow \tilde{\alpha}$

3.3 Description of computer-simulation studies

3.3.1 Methods

The use of SGD and RDA for USCT image reconstruction was evaluated through two-dimensional computer-simulation studies. The studies can be grouped into two parts: (1) studies based on estimation of the SOS distribution for a numerical breast phantom (shown in Fig. 3.2a) and (2) a bias-variance analysis based on a low-contrast phantom with two homogeneous bars (shown in Fig. 3.2b). For both studies, the same measurement geometry, excitation pulse, and numerical simulation methods were employed.

Measurement Geometry

The geometry of the measurement system was chosen to match an existing USCT imaging system [27, 28, 64]. It consisted of a circular transducer array with a radius of 110 mm and 256 evenly distributed elements. The pressure wavefield data were simulated for 256 views using the first-order k-space method as described in Section 2.1.4 [103, 104, 109]. For each view, one transducer served as the emitter and the pressure was recorded by all transducers. All transducers were modeled as ideal point emitters and receivers. Within the ring array was a circular field-of-view over which the SOS distribution was to be estimated (see Fig. 3.1 for a schematic of this measurement geometry).

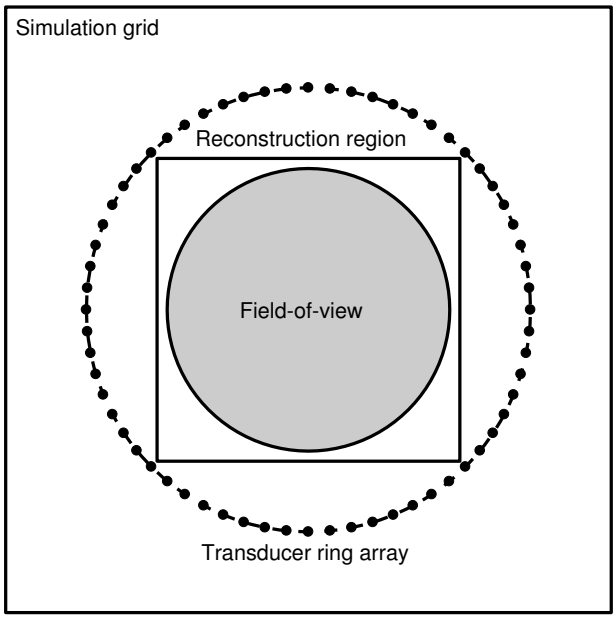


Figure 3.1: A schematic of the measurement geometry. The measurement system consists of a circular ring array of ultrasonic transducers. These transducers are located in a larger rectangular simulation grid, over which the acoustic wave equation is solved. Within the ring array is a smaller rectangular region representing the reconstructed image. The estimated SOS distribution is calculated within the gray circular field-of-view within that region. ()

Excitation pulse

The acoustic excitation pulse was given by

$$s(t) = \exp\left(-\frac{(t-t_c)^2}{2\sigma^2}\right)\sin(2\pi f_c t), \tag{3.11}$$

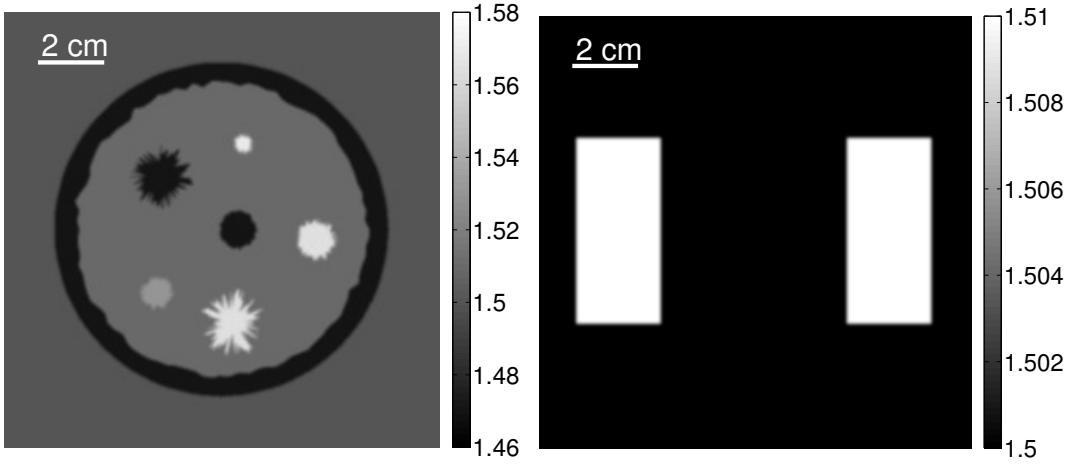
where $f_c = 0.8$ MHz is the central frequency, and $t_c = 3.2$ μs and $\sigma = 0.75$ μs are the center and width of a Gaussian window, respectively. This corresponds to roughly three cycles. Nearest neighbor interpolation was employed to place the transducers on the discrete simulation grid. As a result, the source term for the m-th view is given simply by

$$s_m(\mathbf{r}, t) = s(t)\,\delta(\mathbf{r} - \mathbf{r}_m), \tag{3.12}$$

where $\mathbf{r}_m$ is the location of the pixel nearest to the emitter for the m-th view.

Numerical phantoms

A previously developed numerical breast phantom, composed of 8 structures representing adipose tissues, parenchymal breast tissues, cysts, benign tumors, and malignant tumors, was employed to compare the two optimization methods [109]. It had a radius of 49 mm and SOS values ranging from 1.47 mm/μs to 1.57 mm/μs depending on the tissue type (see Fig. 3.2a). The phantom employed for the bias-variance analysis consisted of two low-contrast homogeneous bars (see Fig. 3.2b). The bars were placed far apart in order to minimize the impact of one bar on the other in the reconstructed images.



(a) (b)

Figure 3.2: Sound speed distribution of (a) the numerical breast phantom and (b) the low-contrast two bar phantom employed in the bias-variance analysis, given in units of mm/μs. (© 2017 IEEE)

Simulation of pressure data

In order to avoid inverse crime [20], both the first-order and second-order k-space pseudo-spectral methods were employed to simulate the pressure. When generating the pressure data recorded by each transducer, the acoustic wave equation was solved by use of the first-order k-space pseudo-spectral method discussed in Sections 2.1.2 and 2.1.4 [103]. The calculation domain consisted of a 2048×2048 uniform Cartesian grid with a pixel size of 0.25 mm, resulting in a physical size of 512×512 mm^2. The pressure was simulated at a sampling rate of 20 MHz for 3600 time points. Additive Gaussian white noise, with zero mean and a standard deviation of 5% of the maximum pressure amplitude received by the transducer opposite the emitter for a homogeneous medium, was added to the pressure data.

When reconstructing the SOS images, the operator $\mathbf{H}_{US}(\mathbf{c})$ was computed by use of the second-order k-space pseudo-spectral method [71]. In this case, the calculation domain consisted of a 1024×1024 uniform Cartesian grid with a pixel size of 0.5 mm, corresponding

to a physical size of 512×512 mm^2. The sampling rate and number of time points were reduced to 10 MHz and 1800, respectively. Both numerical wave solvers were implemented using NVIDIA's CUDA platform [1].

3.3.2 Image reconstruction

Reconstructed images of the numerical breast phantom were obtained by use of both SGD and RDA. In order to provide a clear and fair point of comparison of the two methods, the hyperparameters for each method were individually tuned via a grid search. The value of the regularization parameter was tuned in order to minimize the root-mean-square error (RMSE). In cases where a constant step size was employed, the value of this constant step size was tuned in order to achieve the fastest convergence in terms of the RMSE. While in practice it is not possible to tune these parameters according to the RMSE as the true object is typically unknown, by optimizing the parameters in this way, we compare the two optimization methods irrespective of the particular strategies adopted for choosing the hyperparameters. Developing such strategies is challenging and outside the scope of this current study.

For simplicity of presentation, the results are broken up into three parts. First, results are shown for SGD. These results are meant to serve as a reference. Second, the properties of the RDA algorithm are investigated. Third, the results for SGD and RDA are directly compared. In addition to comparing the final images in terms of RMSE, the convergence rates of the two approaches are also compared. The convergence rates are evaluated in terms of the number of iterations and the number of times that the wave equation must be solved in order to reach a certain accuracy. This later measure, which we refer to as the number of

wave solver runs, is a good proxy for overall computational cost of the method. Use of a line search increases the number of wave solver runs for each iteration relative to use of a constant step size. Every step size considered as part of the line search will add one additional wave solver run.

3.3.3 Bias-variance analysis

A bias-variance analysis was performed to compare the statistical properties of the images produced by SGD and RDA. The measured pressure data were generated using the same procedure described in Section 3.3.1. Then, the measured data were corrupted by five different noise realizations, each with 5% noise. For each noise realization, images were reconstructed for six different regularization parameter values by use of SGD with a constant step size and the unweighted RDA method. Due to the long image reconstruction times (approx. 1 hr for 250 iterations), it was not feasible to reconstruct images for a large number of noise realizations. Instead, each reconstructed image was divided into 20 identical regions (10 per bar), which were treated as independent samples for the purposes of this analysis. The corresponding pixels in these regions were treated as having arisen from additional noise realizations, resulting in a total of 100 samples per regularization parameter value.

Let each pixel in the reconstructed images be treated as a random variable $\hat{c}_i$ (for the i-th pixel), whose true value in the original phantom is c_i. Further, let $\widehat{C}_i$ denote the set of values for the i-th pixel for the five noise realizations. Then, an augmented set $\widetilde{C}_i$ can be created such that

$$\widetilde{C}_i = \bigcup_{j=1}^{N_c} \widehat{C}_{\mathcal{I}_i(j)}, \tag{3.13}$$

where N_c is the total number of regions (20) and $\mathcal{I}_i$ is an iterator that gives the indices of all pixels (across regions) that correspond to the i-th pixel.

The bias for each pixel within a region was calculated by averaging the 100 samples and computing the difference between the average value and the corresponding value in the true phantom:

$$\text{Bias}_i = \frac{1}{N_s} \sum_{\hat{c} \in \widetilde{C}_i} \hat{c} - c_i, \tag{3.14}$$

where N_s is the total number of samples. As a summary measure, the ℓ_2-norm of the bias values across pixels was calculated. Similarly, the sample variance of each pixel across all samples was computed as

$$\text{Var}_i = \frac{1}{N_s - 1} \sum_{\hat{c} \in \widetilde{C}_i} \left(\hat{c} - \frac{1}{N_s} \sum_{\hat{c} \in \widetilde{C}_i} \hat{c} \right)^2. \tag{3.15}$$

As a summary measure, the average variance for the pixels was computed. While corresponding pixels in different regions may not have the same expected values and variances, the above bias and variance measures still provide insight into the ability of the two reconstruction algorithms to mitigate noise.

3.4 Results of computer-simulation studies

3.4.1 Images reconstructed by SGD

First, the regularization parameter value for SGD that results in the most accurate reconstructed image, as quantified by the RMSE, was determined. As seen in Fig. 3.3, a regularization parameter value of 5×10^{-4} results in the most accurate reconstructed image. As such, this value will be taken as the optimal value for SGD-based USCT image reconstruction and will be employed when making comparisons with the results obtained by use of the RDA method.

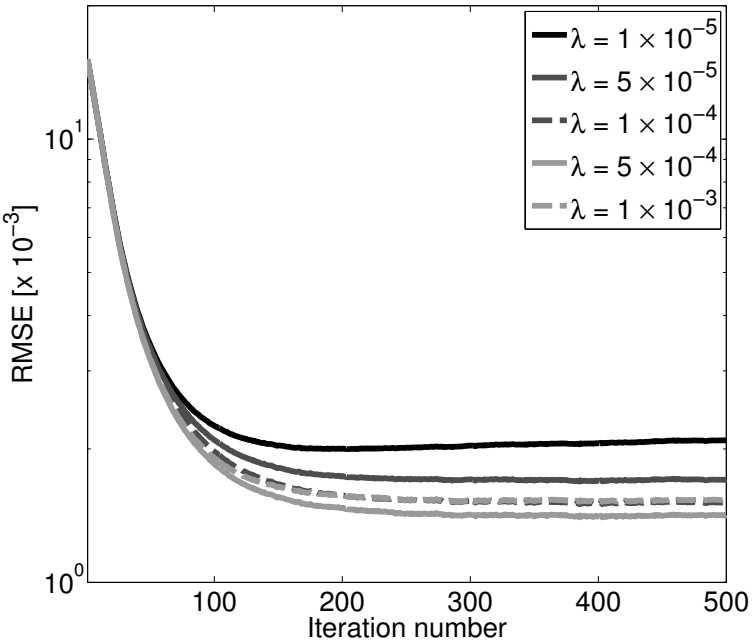
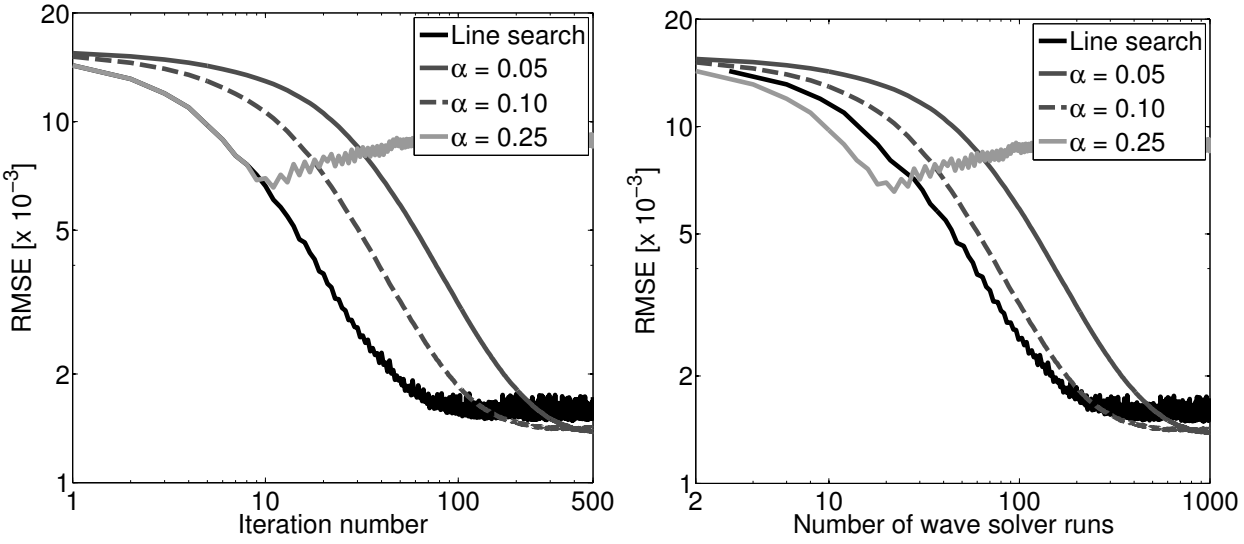


Figure 3.3: Plot of RMSE versus the number of iterations for images reconstructed by use of SGD with a constant step size of 0.1 for several regularization parameter values. (© 2017 IEEE)

Next, the optimal constant step size for SGD was estimated by sweeping over a range of step sizes. These results were also compared with use of a line search method. As seen in Fig. 3.4,

the optimization algorithm will diverge when the constant step size is too large. However, when the step size is too small, the convergence of the optimization algorithm will be slow. Use of a line search method produces the fastest initial convergence, both in terms of the number of iterations and in terms of the number of wave solver runs. In addition, use of a line search removes the need to determine an appropriate constant step size, a time-consuming task which is often accomplished through trial-and-error. From Fig. 3.4, it can be seen that a constant step size of 0.1 produces the fastest convergence rate among all the constant step size results while still resulting in an accurate reconstructed image.

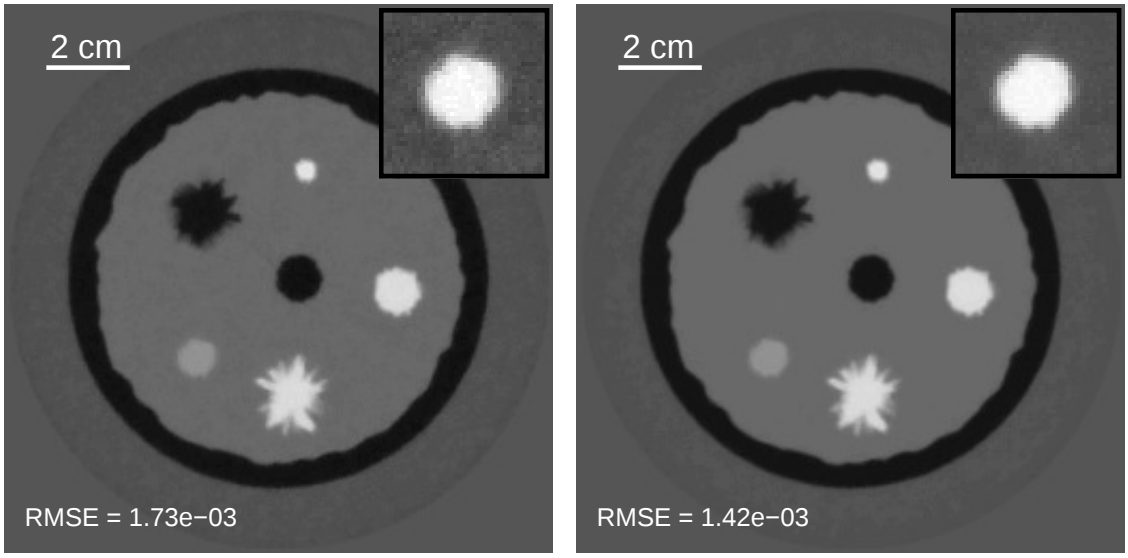


(a) (b)

Figure 3.4: Plot of RMSE versus (a) the number of iterations and (b) the number of wave equation solver runs for images reconstructed by use of SGD with a line search and with several constant step size values for a regularization parameter value of 5×10^{-4}. (© 2017 IEEE)

However, there are some downsides to employing a line search method with SGD. As can be seen in Fig. 3.4, use of a line search results in oscillations in the RMSE of the reconstructed image, while use of a constant step size produces a smoother convergence curve with fewer jumps. In addition, the final RMSE is lower for the constant step size method

(RMSE = 1.42×10^{-3}) than for the line search method (RMSE = 1.73×10^{-3}). These differences are also reflected in the apparent image quality of the reconstructed images (see Fig. 3.5).



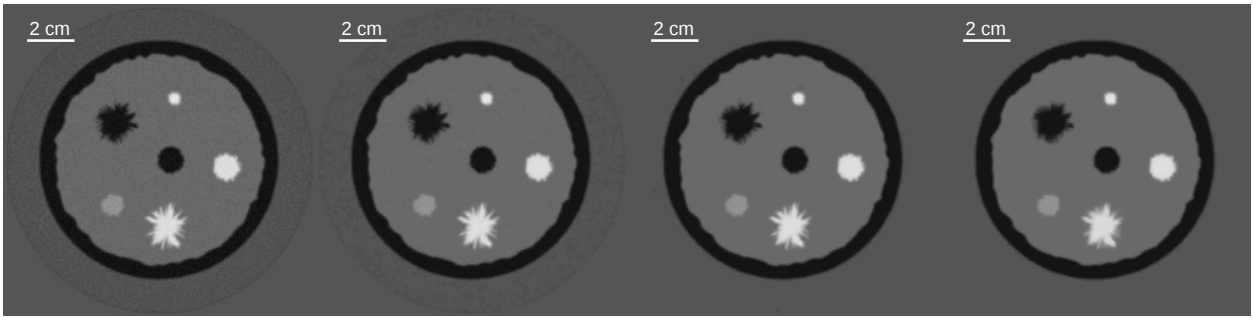
(a) (b)

Figure 3.5: Images reconstructed by use of SGD using (a) a line search and (b) a constant step size of 0.1, shown after 500 iterations for a regularization parameter value of 5×10^{-4}. The insets in the upper right corners of each image correspond to a zoomed-in image of the inclusion located at about 7 o'clock. The larger images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs, while the insets are shown in a grayscale window of $[1.50, 1.53]$ mm/μs. (© 2017 IEEE)

3.4.2 Images reconstructed by RDA

The regularization parameter value that resulted in the most accurate reconstructed image was also determined for the RDA method. Reconstructed images for different regularization parameter values, shown in Fig. 3.6, suggest that TV regularization behaves largely as expected for this new optimization method. The impact of the regularization term increases

smoothly and continuously as the the value of the regularization parameter is increased. When the regularization parameter is small, a noisy background texture to the reconstructed images can be detected. When the regularization parameter is large, a smoother background is observed. From Fig. 3.7a, it can be seen that a regularization parameter value of 1×10^{-4}

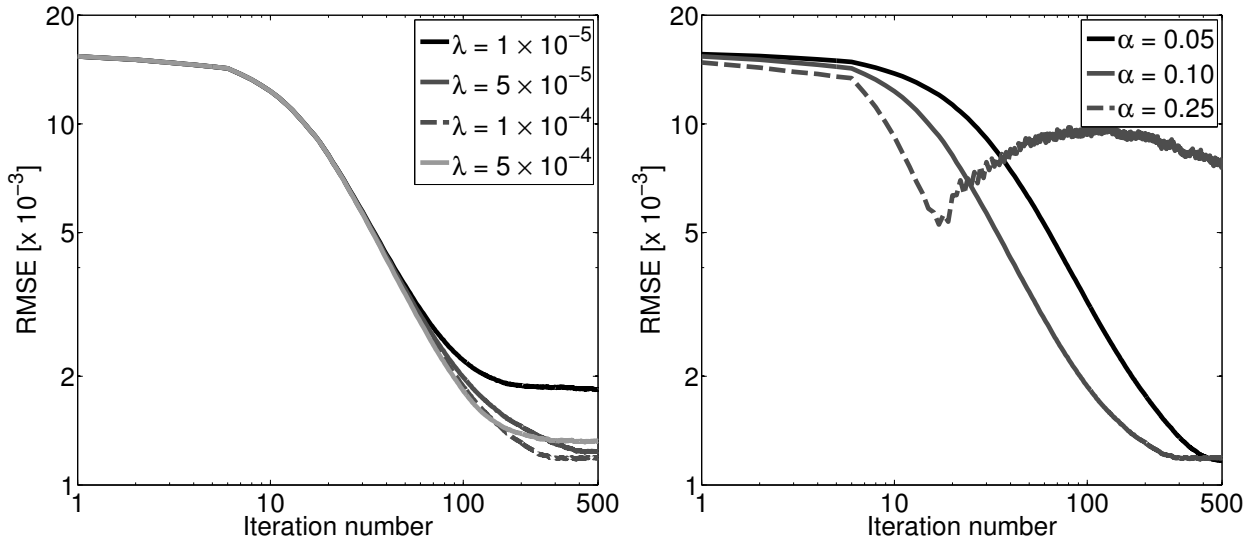


(a) (b) (c) (d)

Figure 3.6: Images reconstructed by the unweighted RDA method with a fixed step size of 0.1 for regularization parameter values of (a) 1×10^{-5}, (b) 5×10^{-5}, (c) 1×10^{-4}, and (d) 5×10^{-4}, shown after 300 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE)

results in the most accurate reconstructed image for the RDA method. This is smaller than the value obtained for SGD, emphasizing the importance of individually tuning hyperparameters for each optimization method. Next, the optimal constant step size (or, equivalently, value of γ in line 9 of Algorithm 2) was determined. From Fig. 3.7b, the optimal constant step size value is 0.1, the same value obtained for SGD.

The weighted RDA method, in which the gradients at different iterations are given unequal weights, can be used to achieve faster convergence compared with the unweighted case. To confirm that this weighting strategy did not have a large impact on the regularization, images were reconstructed for several regularization parameter values (see Fig. 3.8). The weighting strategy does not appear to have had an outsized effect on the regularization performance. As with the unweighted case, a regularization parameter value of 1×10^{-4} results in the



(a) (b)

Figure 3.7: Plot of RMSE versus the number of iterations for (a) several regularization parameter values and a fixed step size of 0.1 and (b) several constant step size values and a fixed regularization parameter value of 1×10^{-4} for images reconstructed by use of the unweighted RDA method. (© 2017 IEEE)

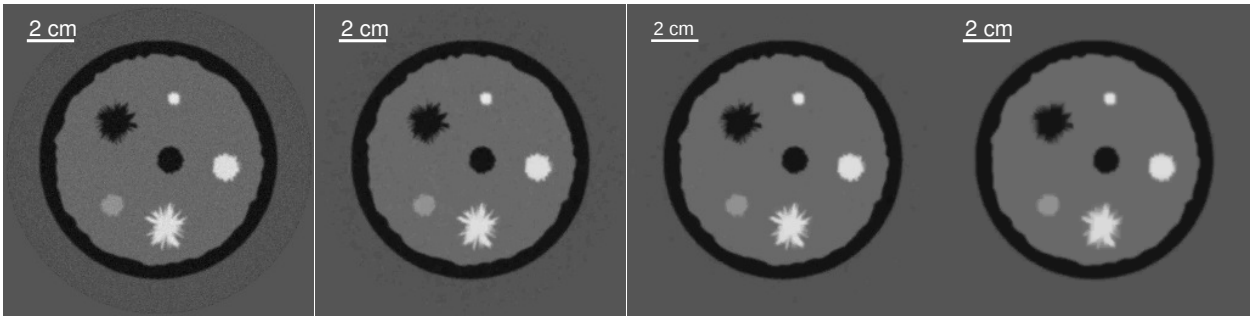


Figure 3.8: Images reconstructed by the weighted RDA method for regularization parameter values of (a) 1×10^{-5}, (b) 5×10^{-5}, (c) 1×10^{-4}, and (d) 5×10^{-4}, shown after 300 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE)

smallest RMSE. While the final image obtained after many iterations is largely unchanged by the weighting strategy, the accuracy of reconstructed images obtained at early iterations is greatly improved. As seen in Fig. 3.9, the RMSEs and the apparent visual quality of the reconstructed images after 20, 50, and even 100 iterations are improved by use of the weighted RDA method. This improved accuracy can also be seen in the profiles through

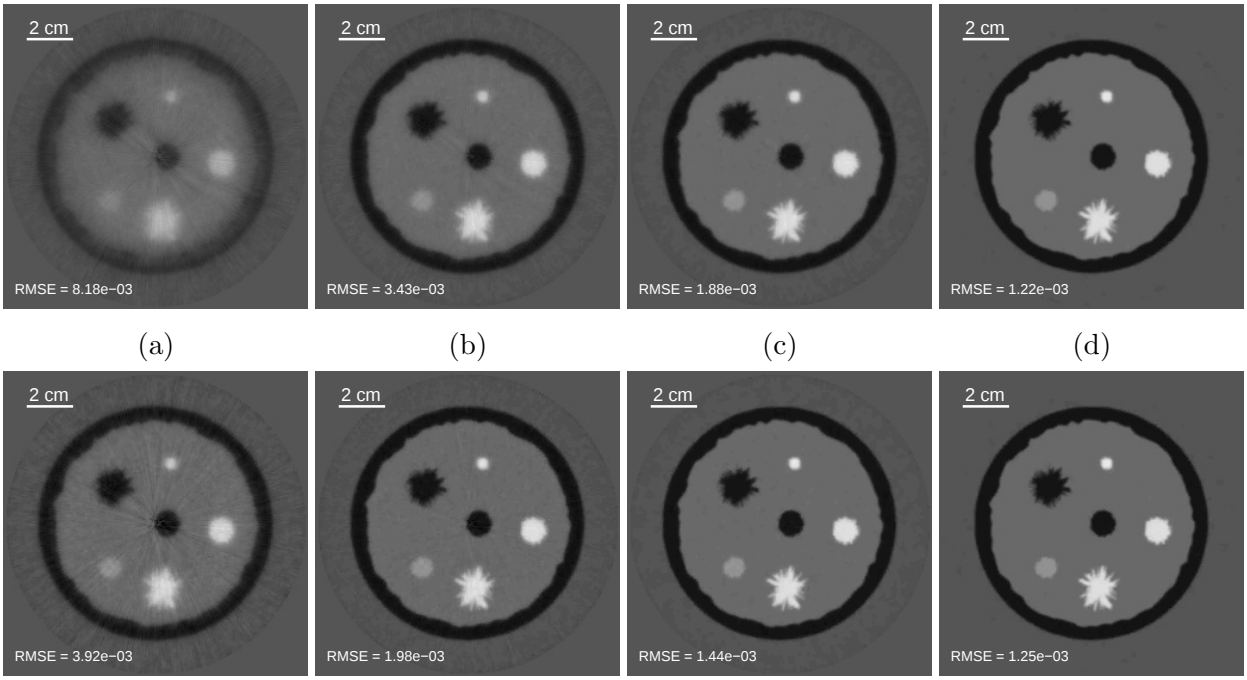
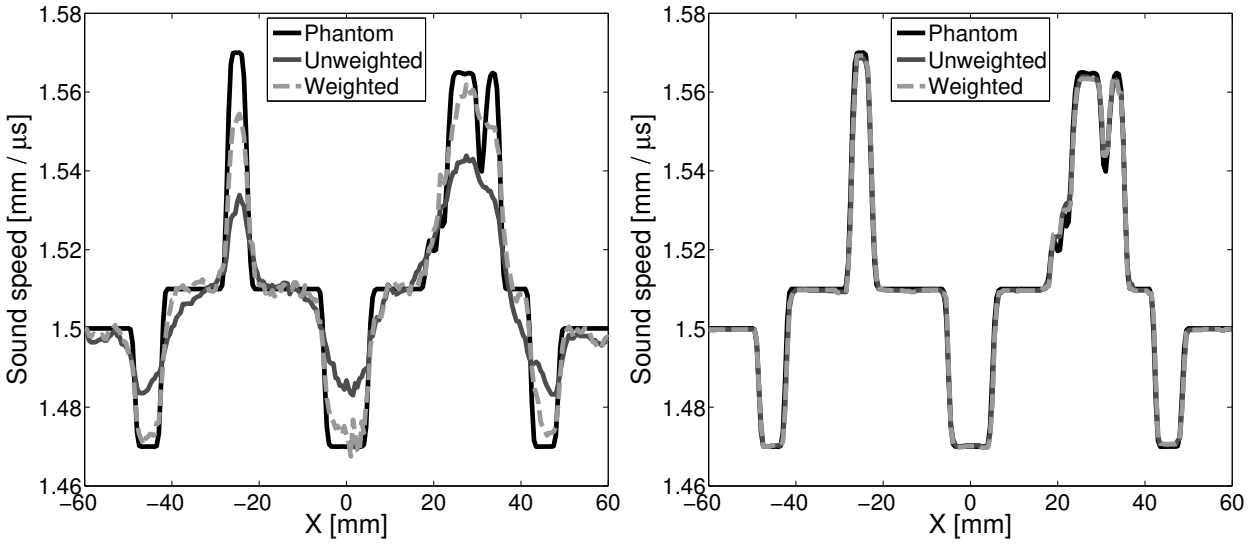


Figure 3.9: Images reconstructed by use of the unweighted RDA method with a fixed step size of 0.1 after (a) 20, (b) 50, (c) 100, and (d) 250 iterations. Images reconstructed by use of the weighted RDA method after (e) 20, (f) 50, (g) 100, and (h) 250 iterations. All results are shown for a regularization parameter value of 1×10^{-4} and in a grayscale window of $[1.47, 1.58]$ mm/μs. The RMSEs for each reconstructed image are displayed in the bottom left of each subfigure. (© 2017 IEEE)

the reconstructed images shown in Fig. 3.10. As shown in Fig. 3.11, this improvement is maintained even when the convergence of the reconstruction methods is viewed in terms of the number of wave solver runs as opposed to the number of iterations. Eventually,



(a) (b)

Figure 3.10: (a) Profiles through y = -6.5 mm for reconstructed images obtained by use of the weighted RDA method and the unweighted RDA method with a fixed step size of 0.1, shown after (a) 20 iterations and (b) 250 iterations. (© 2017 IEEE)

after approximately 250 wave solver runs (or 250 iterations for the unweighted method), the weighted and unweighted approaches produce images of similar accuracy.

3.4.3 Images reconstructed by SGD and RDA

In the previous sections, the SGD and RDA methods were evaluated individually. Now, the images produced by use of the two methods are compared directly. Images reconstructed by four different optimization strategies are shown in Fig. 3.12: (1) SGD with a constant step size, (2) unweighted RDA, (3) SGD with a line search, and (4) weighted RDA. The RMSE of each reconstructed image is noted in the bottom left. Together with the apparent visual quality of the images, this measure demonstrates that the initial convergence rates of SGD with a line search and the weighted RDA method are much faster than that of either SGD

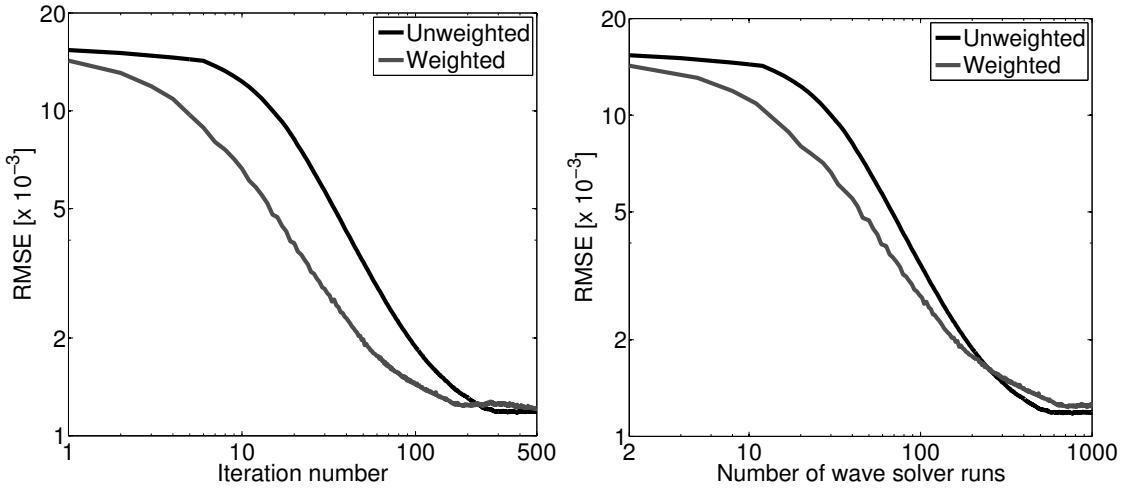


Figure 3.11: Plot of RMSE vs. (a) the number of iterations and (b) the number of wave solver runs for the weighted and unweighted RDA methods. (© 2017 IEEE)

with a constant step size or the unweighted RDA method. However, the two RDA methods produce reconstructed images with higher accuracy at later iterations when compared with the SGD-based methods. Specifically, although use of a line search limited the final accuracy of images produced by SGD, the accuracy of the image reconstructed by the weighted RDA method is better than that obtained by SGD with a constant step size. This demonstrates that the weighted RDA method can provide both fast convergence and more accurate images than using the SGD method.

As shown in Fig. 3.13, the improved accuracy of the weighted RDA method compared with SGD with a line search is reflected in the profiles through the reconstructed images. The profile for the image obtained by use of SGD is noticeably noisier than that obtained by use of RDA. This may indicate that the RDA method is more effective at mitigating noise than SGD. In Section 3.4.4, this potential benefit will be considered more closely through the use of a bias-variance analysis.

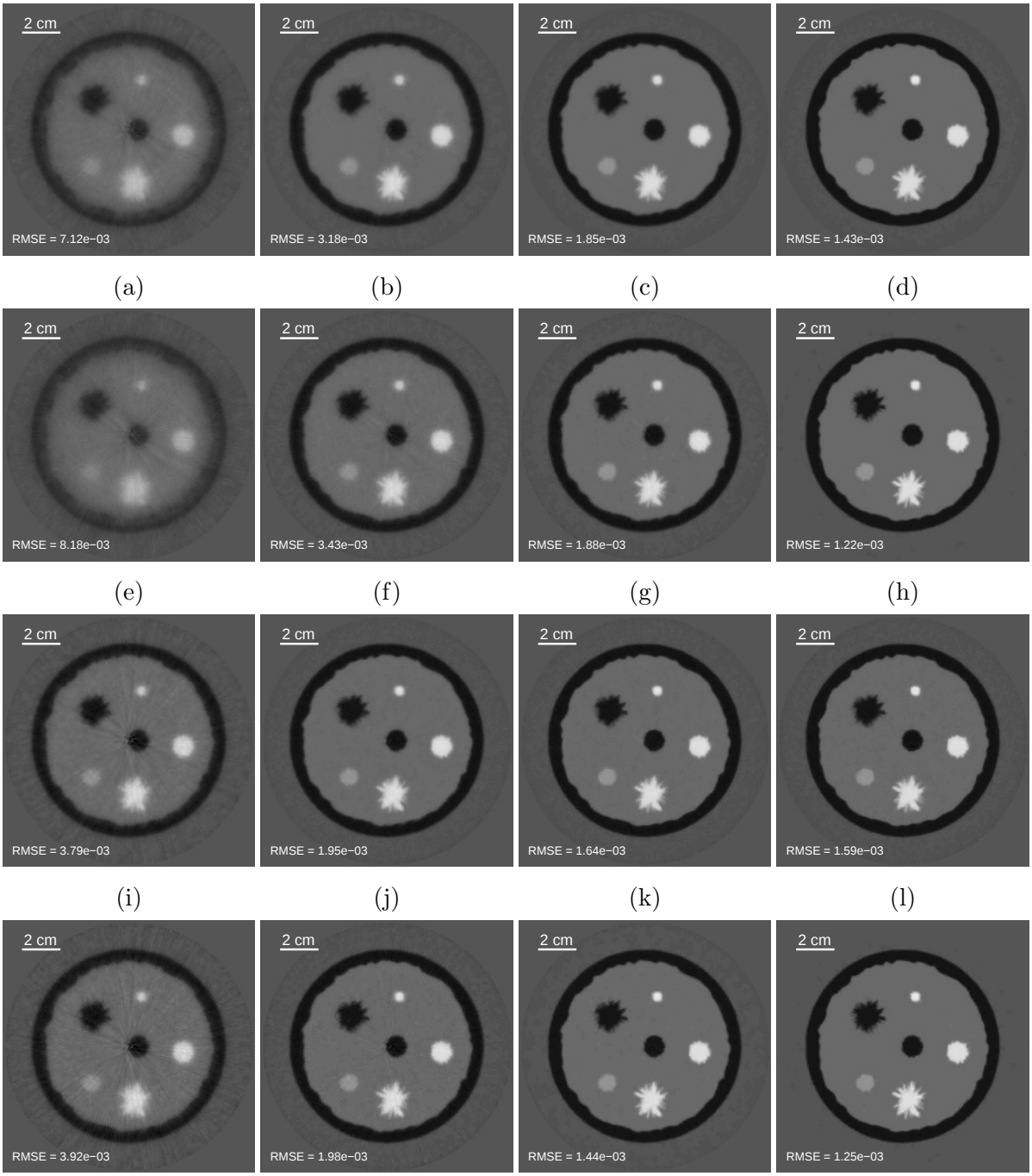


Figure 3.12: Images reconstructed by (Row 1) SGD with a constant step size of 0.1 and $\lambda = 5\times10^{-4}$; (Row 2) unweighted RDA with a fixed step size of 0.1 and $\lambda = 1\times10^{-4}$; (Row 3) SGD with a line search and $\lambda = 5 \times 10^{-4}$; and (Row 4) weighted RDA with $\lambda = 1 \times 10^{-4}$. The columns correspond to the images obtained, from left to right, after 20, 50, 100, and 250 iterations. All images are shown in a grayscale window of $[1.47, 1.58]$ mm/μs. (© 2017 IEEE)

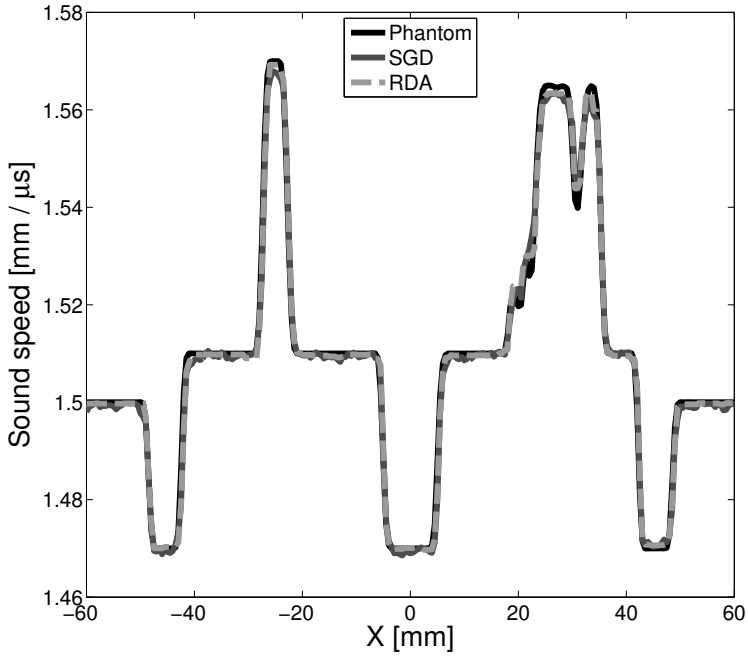
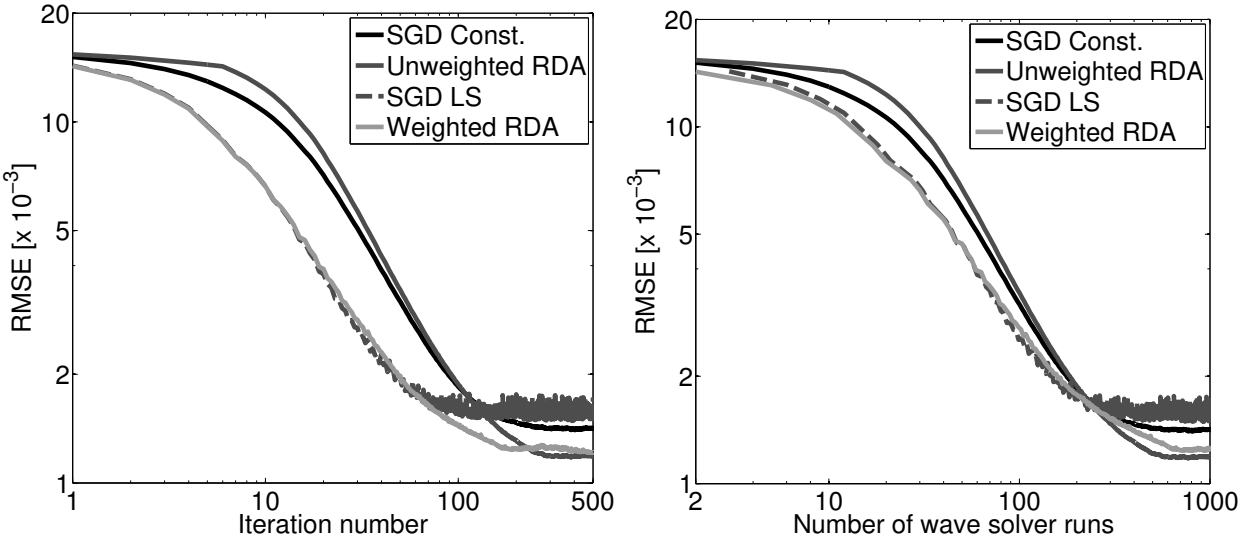


Figure 3.13: Profiles through y = -6.5 mm for images reconstructed by the use of SGD with a line search and weighted RDA. (© 2017 IEEE)

The benefits provided by the RDA method are further confirmed by the plots of the convergence rates, shown in Fig. 3.14. While SGD with a line search has a fast initial convergence rate, it ultimately results in a less accurate final image. In addition, the estimates of the object provided by SGD with a line search also exhibit a high variance, even at later iterations. This is likely due to overfitting, caused by the fact that the line search only evaluates the cost function for a single realization of the encoding vector. As a result, the line search will overestimate how much moving in the given search direction will reduce the overall cost function. As a result, the line search will tend to chose a larger step size that effectively minimizes the cost function evaluated for that encoding vector, but which increases the cost function when all, or a large number, of encoding vectors are considered. This behavior is not observed for the weighted RDA method. For the RDA method, the search direction is

given by a weighted average of the gradient estimates for all past encoding vector realizations. Consequently, it does not overfit the cost function evaluated for a single realization of the encoding vector. Thus, the high variance of the object estimates is eliminated while the computational cost of the line search for the RDA method is the same as for SGD.



(a) (b)

Figure 3.14: Plot of RMSE versus (a) the number of iterations and (b) the number of wave solver runs for SGD with a line search, SGD with a constant step size of 0.1, unweighted RDA with a step size of 0.1, and weighted RDA. (© 2017 IEEE)

3.4.4 Bias-variance analysis

Previously, it was observed that the RDA method could produce reconstructed images with lower RMSEs than SGD. One possible explanation for this observation is that the regularization term under the RDA method is able to more effectively mitigate noise. The hypothesis is consistent with the idea that grouping together the stochastic data fidelity term and the deterministic regularization term may not be optimal. To evaluate this idea, it is necessary

to confirm that the observed differences between the two methods are not simply due to superior hyperparameter tuning in one of the two cases.

Bias-variance curves, which depict the inherent trade-off between noise mitigation and close agreement with the measured data, offer one solution for systematically evaluating the two methods. As described in Section 2.2.3, an estimate of the SOS distribution is obtained by solving an optimization problem consisting of two terms, the data fidelity term and the regularization term. The relative weight of these terms is controlled by a scalar regularization parameter. For TV regularization, noise can be more severely suppressed by increasing the value of the regularization parameter at the possible expense of reduced resolution or other forms of bias.

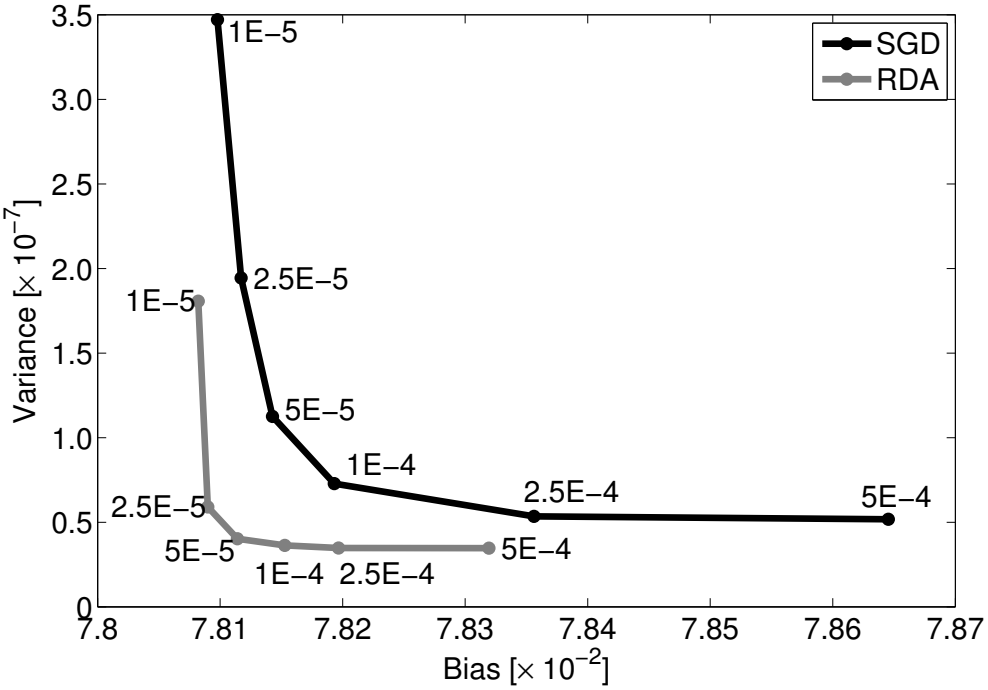
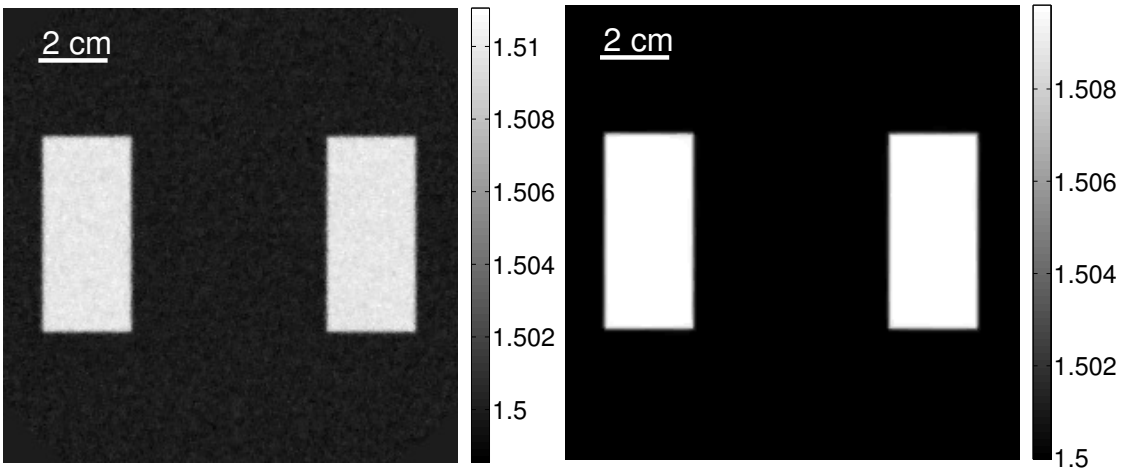


Figure 3.15: Bias-variance curve for SGD with constant step size and the unweighted RDA method. The corresponding regularization parameter values are given by each point. (© 2017 IEEE)

Bias-variance curves were generated for SGD with a constant step size and the unweighted RDA method by reconstructing a collection of images for six different regularization parameter values. As seen in Fig. 3.15, the RDA method consistently produces lower variance images (less noisy) for a given level of bias. This difference can also be seen in the reconstructed images shown in Fig. 3.16. In that figure, the reconstructed images have the same bias level, but the image reconstructed by use of SGD with a constant step size is noticeably noisier than the image obtained by use of the unweighted RDA method.



(a) (b)

Figure 3.16: Example reconstructed images from bias-variance analysis. (a) Image reconstructed by SGD with a regularization parameter value of 5×10^{-5}. (b) Image reconstructed by RDA with a regularization parameter value of 1×10^{-4}. The two images have approximately the same bias. Both images are shown in their full dynamic ranges. The SOS values are given in units of mm/μs. (© 2017 IEEE)

3.5 Description of clinical studies

The utility of the proposed approach was also evaluated through clinical studies. The clinical data were acquired previously by Delphinus Medical Technologies and the Karmanos Cancer

Institute by use of the SoftVue USCT scanner [28]. The imaging system consists of a ring array with a radius of 110 mm and 2048 evenly-distributed transducers. The ultrasonic transducers had a central frequency of 2.75 MHz with a pitch of 0.34 mm. Each element was elevationally focused to isolate a 3-mm-thick slice of the breast. Additional information regarding the measurement system and clinical studies can be found in [27, 28].

Every other transducer element served as an emitter, and the resulting pressure waves were then measured by the same set of 1024 transducers. The pressure data were recorded by the ultrasonic transducers at a sampling rate of 12 MHz for 2112 time points, corresponding to approximately 176 μs. As a calibration step, this measurement process was repeated without the object. By manual inspection, 48 transducers were identified as having low sensitivity or poor performance. The data from these transducers were discarded, resulting in measurements from 976 transducers. The pressure data were upsampled to a sampling rate of 20 MHz by use of linear interpolation in order to avoid the introduction of numerical errors by the wave solver [71]. After upsampling, the number of samples in each time trace was 3500. Additionally, each signal was filtered by a Butterworth bandpass filter with cutoff frequencies of 0.5 and 1.0 MHz. The shape of the excitation pulse was estimated from the calibration data taken without the object using the method described in [109].

An initial estimate of the SOS distribution was reconstructed by use of an adjoint state method (see Fig. 3.17) [6]. This estimate was also used to generate a set of synthetic measured data. As detailed in [109], measurements near the emitter may not contribute positively to the reconstructed image due to a variety of factors, including mechanical cross-talk, model mismatch, and measurement noise. To mitigate the impact of these effects, the measurements near the emitter were replaced with synthetic data. Only the 512 measurements from transducers opposite the emitter were kept.

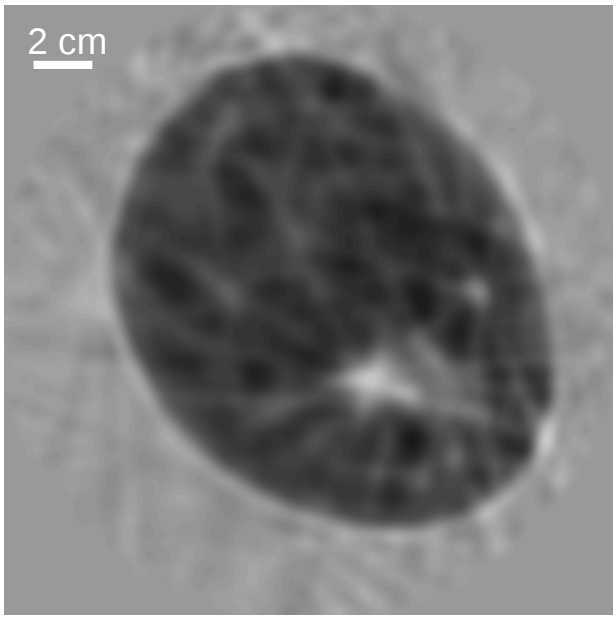


Figure 3.17: Initial estimate of the object reconstructed by use of an adjoint state method described in [6]. (© 2017 IEEE)

The images were reconstructed by solving Eqn. 2.39, where the operator $\mathbf{H}_{US}(\mathbf{c})$ was calculated by use of the second-order k-space pseudo-spectral wave solver as described in Section 3.3.1 [71]. The calculation domain consisted of a 2560×2560 Cartesian grid with a pixel size of 0.2 mm, corresponding to a physical size of 512×512 mm^2. The SOS was updated within a field-of-view of radius 105 mm. The images were reconstructed using a NVIDIA Tesla K40 GPU.

While image quality is most objectively assessed using task-based methods of image quality [10], such studies are often a substantial undertaking. As a result, the contrast-to-noise (CNR) ratio was employed as a proxy for the detectability of the tumor. The CNR of the reconstructed images was calculated by identifying three regions: (1) the tumor, (2) parenchymal breast tissue, and (3) the water bath. The tumor was identified by visual inspected and manually segmented. Circular regions of a similar size corresponding to the

parenchymal tissue and the water bath were also identified. The contrast was calculated by comparing the tumor and parenchymal tissue regions. The noise, however, was calculated based on the water bath, which was assumed to be homogeneous, to avoid mis-attributing any real variations within the parenchymal tissue to noise. With this, the CNR was calculated as

$$CNR = \frac{\bar{c}_t - \bar{c}_p}{\sigma_n}, \tag{3.16}$$

where $\bar{c}_t$ is the average SOS of the tumor, $\bar{c}_p$ is the average SOS over a comparably sized region of the parenchymal tissue, and σ_n is the standard deviation over a comparably sized region of the water bath.

3.6 Results of clinical studies

Images were reconstructed for several different regularization parameter values for both SGD with a constant step size and the weighted RDA methods. The CNRs of each image are indicated in the bottom left corner of the image. As seen in Fig. 3.18, the weighted RDA method consistently produces reconstructed images with higher CNRs than SGD with a constant step size. This is observed across a range of regularization parameter values. Further, the CNR obtained by SGD is lower even when the regularization parameter value is tuned to maximize the CNR. This improvement in the CNR for the RDA method is consistent with the superior noise mitigation performance observed in the computer-simulation studies. While the CNRs of all the images shown in Fig. 3.18 are quite high, the improvement in the CNR could be more impactful for cases where the tumors are smaller or have lower contrast.

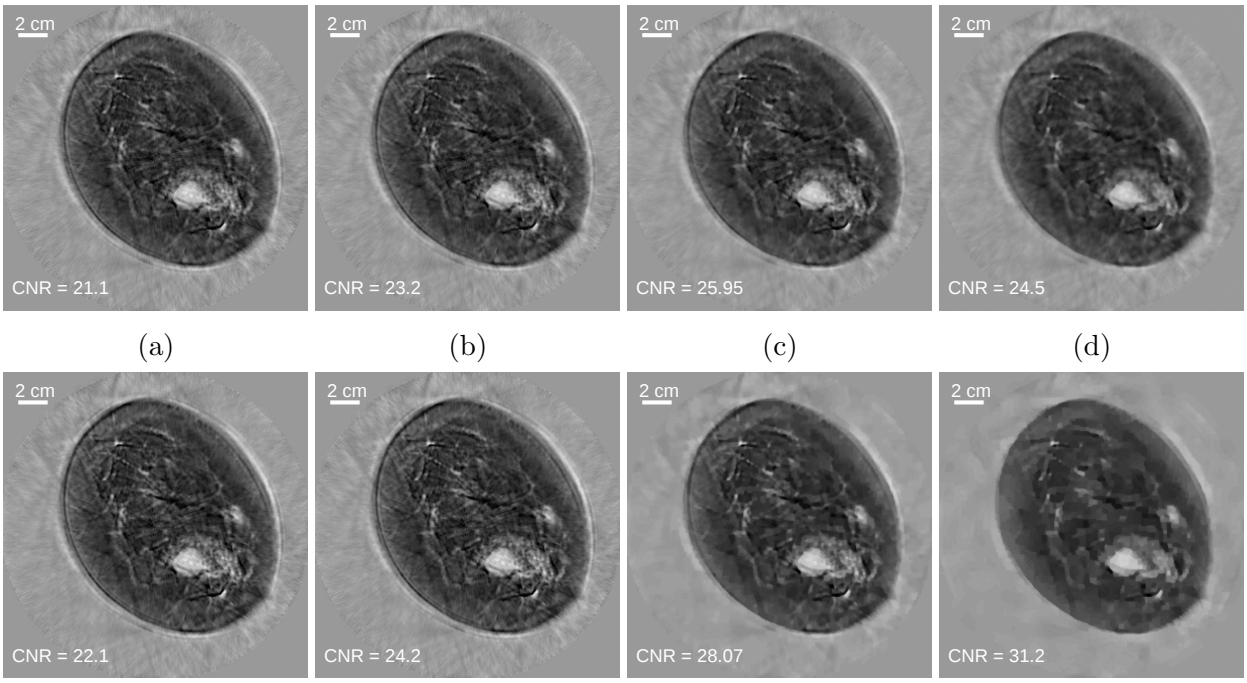


Figure 3.18: (Top row) Images reconstructed by use of SGD with a constant step size of 2.5×10^5 and regularization parameter values of (a) 1×10^{-10}, (b) 3×10^{-10}, (c) 1×10^{-9}, and (d) 3×10^{-9}. (Bottom row) Images reconstructed by use of the weighted RDA method with regularization parameter values of (e) 1×10^{-10}, (f) 3×10^{-10}, (g) 1×10^{-9}, and (h) 3×10^{-9}. Images are shown after 100 iterations and in a grayscale window of $[1.38, 1.60]$ mm/μs. (© 2017 IEEE)

To compare the differences in the convergence rates of the two optimization methods, reconstructed images for a series of iteration numbers are shown in Fig. 3.19. Since a non-constant initial guess was provided, the differences in the convergence rates of SGD with a constant step size and the weighted RDA method are less pronounced than observed in the computer-simulation studies. Still, the weighted RDA method produces a higher CNR at each iteration even though the regularization parameter value was chosen to maximize the CNR for the SGD case. The difference between the CNRs of the two methods grows over the first 50 iterations before decreasing at later iterations. This might suggest that the weighted RDA method is able to provide some initial improvement in the convergence rate as was observed with the computer-simulation studies.

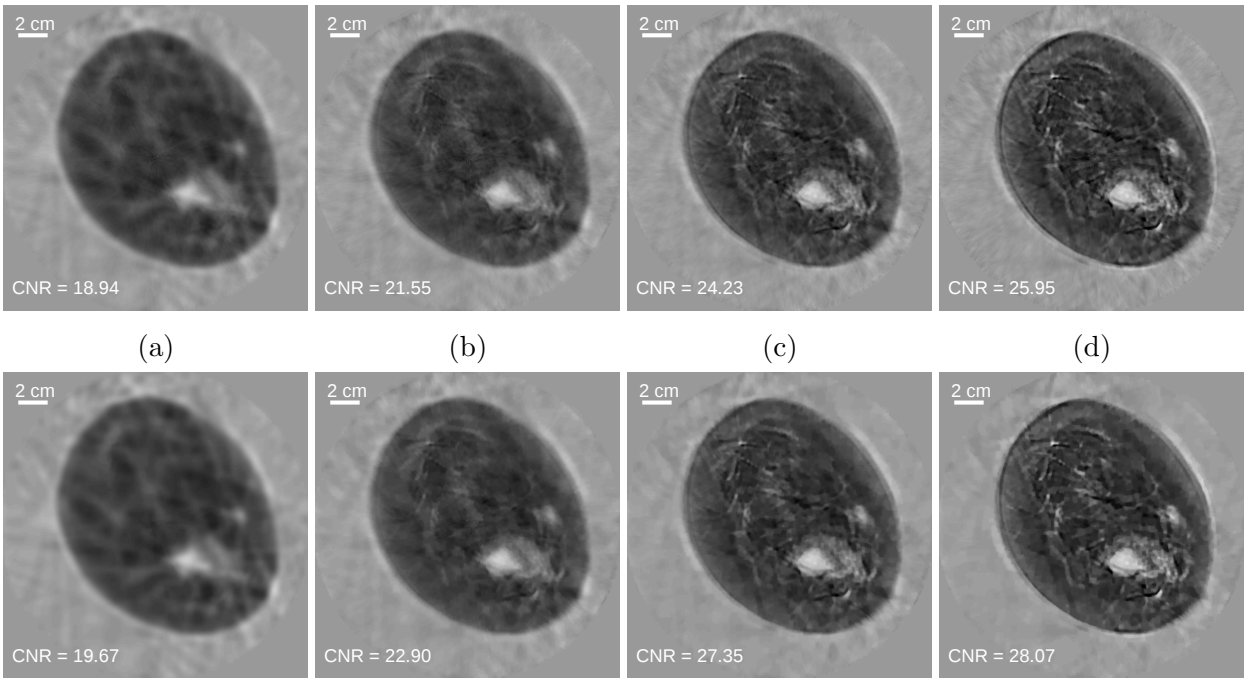
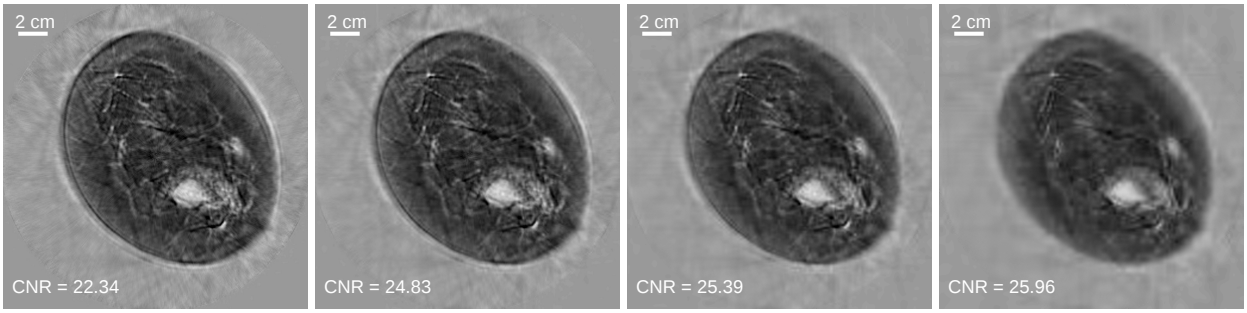


Figure 3.19: (Top row) Images reconstructed by use of SGD with a constant step size of 2.5×10^5 after (a) 5, (b) 20, (c) 50, and (d) 100 iterations with a regularization parameter value of 1×10^{-9}. (Bottom row) Images reconstructed by use of weighted RDA after (e) 5, (f) 20, (g) 50, and (h) 100 iterations with a regularization parameter value of 1×10^{-9}. All images are shown in a grayscale window of [1.38, 1.60] mm/μs. (© 2017 IEEE)

Another key benefit of the RDA method is its ability to allow natural incorporation of non-smooth penalties. This may allow the optimization problem be tailored for a given image reconstruction task or to exploit known properties about the object to be reconstructed. While the determination of an optimal choice of regularization function is outside the scope of this dissertation, to demonstrate the flexibility of RDA method, a second non-smooth regularization function was considered. The regularization function was chosen to be

$$\mathcal{R}(\mathbf{c}) = \|\mathbf{\Phi c}\|_1, \tag{3.17}$$

where $\mathbf{\Phi}$ is the 2-D wavelet transform and the mother wavelet was the 12-tap Daubechies wavelet [69]. The wavelet transform was computed by use of the GNU Scientific Library [37]. Images reconstructed with several regularization parameter values are shown (see Fig. 3.20).



(a) (b) (c) (d)

Figure 3.20: Images reconstructed by use of the weighted RDA method with a wavelet-based penalty and regularization parameter values of (a) 3×10^{-10}, (b) 1×10^{-9}, (c) 3×10^{-9}, and (d) 1×10^{-8}. Images are shown after 100 iterations and in a grayscale window of $[1.38, 1.60]$ mm/μs. (© 2017 IEEE)

3.7 Conclusions

Waveform inversion with source encoding can produce high-resolution SOS images while reducing the computational burden that have hindered other time-domain waveform inversion approaches. Using this method, estimates of the SOS distribution can be obtained by minimizing an objective function consisting of a stochastic data fidelity term and a deterministic regularization term. The RDA method offers several advantages when solving this optimization problem compared with use of SGD. First, it exploits knowledge of the structure of the cost function, resulting in more effective regularization. In the case of the TV semi-norm, noise is more effectively reduced while preserving the accuracy and contrast of the reconstructed images. Second, it does not assume that all terms in the regularization function are differentiable, allowing natural incorporation of non-smooth penalties, such as the TV semi-norm and the ℓ_1-norm of the wavelet transform. Third, it exploits information from past iterations to determine the search direction. This allows the method to employ a line search while avoiding overfitting the cost function evaluated for a particular realization of the encoding vector. This allows a fast initial convergence rate without sacrificing the accuracy of the final image. These benefits were demonstrated through computer-simulation studies involving a numerical breast phantom, generation of a bias-variance curve, and experimental studies involving clinical data.

Still, several opportunities for further improvement exist. The assumed forward model ignores a number of important factors that could lead to artifacts in the reconstructed images. In particular, the model ignores mass density variations, acoustic attenuation, dispersion, and out-of-plane scattering. Since the assumed imaging model is 2-D, scattering out of the plane defined by the transducer ring array is not considered. It also treats the transducers

as ideal point detectors and emitters, ignoring their spatial and electro-acoustic impulse responses. Additional investigation of the numerical properties of this approach remains an interesting topic for future study. For example, the frequency content of the excitation pulse and the strength of the acoustic heterogeneities have a sizable impact on the reconstructed images [109]. Comparison with other USCT image reconstruction methods is also needed, e.g. [2, 38, 94, 113].

Chapter 4

Joint reconstruction of the speed of sound and initial pressure distributions from PACT data alone

4.1 Overview

Traditional image reconstruction methods for estimating the initial pressure distribution in PACT assume that the medium is acoustically homogeneous even though this is not true for many biomedical applications [45,88,120]. To mitigate artifacts induced by this assumption, half-time and partial-time image reconstruction methods, which seek to exploit redundant information in data and eliminate parts of the data that are disproportionately affected by acoustic inhomogeneities, have been proposed [3, 4, 88]. Some recent image reconstruction methods have the ability to compensate for SOS variations if the SOS distribution is known [7, 44, 46, 65, 91, 101, 106], but in practice, the SOS distribution is not known. Adjunct imaging data, such as USCT measurements, could be acquired to allow estimation of the SOS distribution (this approach is discussed in detail in Chapter 5) [51, 70, 73, 115], but not

all PACT imaging systems are capable of acquiring the necessary measurements. Thus, there is a need to estimate some approximation of the SOS distribution from PACT measurements alone.

One approach for image reconstruction methods that assume a constant SOS is to tune the SOS according to some image quality metric. For example, Treeby, et al., proposed a method by which the SOS was automatically chosen by maximizing a measure of image sharpness [105]. However, such approaches can be difficult to apply when the assumed SOS is not constant and more SOS values need to be estimated.

Another option is to jointly estimate the SOS and initial pressure distributions from PACT measurements alone [47, 55, 66, 82, 100, 122, 123]. Others have also considered the related problem of joint estimation of the SOS and optical absorption coefficient distributions [25]. However, this approach can be very challenging. For example, Stefanov and Uhlmann considered the joint estimation problem for the linearized wave equation and demonstrated that, in general, the SOS and initial pressure distributions could not both be stably recovered [100]. Huang, et al. made similar observations via computer-simulation studies for the full acoustic wave equation [47].

Previously, Zhang et al. proposed use of a low-dimensional representation of the SOS distribution in order to stabilize the joint reconstruction (JR) problem [123]. In that work, the SOS and initial pressure distributions were estimated by minimizing an objective function consisting of a data fidelity term and a pair of regularization terms by use of a gradient-based alternating minimization approach. The acoustic wave propagation was modeled by use of a generalized Radon transform model, which assumes that the acoustic heterogeneities are relatively weak. In addition, the gradient of the objective function with respect to the parameterized SOS distribution was calculated using a finite-difference-based approach, which

can be computationally expensive when the number of parameters in the SOS distribution is large.

Here, the idea of stabilizing the JR problem by use of a parameterized SOS model is revisited and extended in several key ways. First, the forward model is replaced with a full-wave acoustic model, which does not assume that the acoustic heterogeneities are weak. This allows consideration of bone and other materials, whose acoustic properties differ greatly from the background medium. Second, the adjoint state method is employed to compute the gradients with respect to the initial pressure and SOS distributions. This efficient approach has a computational cost that is independent of the number of parameters in the parameterized SOS model. Third, the optimization method is updated to allow use of constraints and non-smooth regularization terms. Fourth, the feasibility of the approach is demonstrated through application to experimental data.

4.2 Parameterized speed of sound

To constrain and stabilize the JR problem, a low-dimensional parameterized representation of the SOS distribution is considered. Let Q denote the number of unique SOS values in this parameterized representation. Then, the parameterized vector of SOS values $\mathbf{c}_p \in \mathbb{R}^Q$ can be related to the SOS over the whole simulation grid as

$$\mathbf{c} = \mathbf{\Phi}\mathbf{c}_p, \tag{4.1}$$

where $\mathbf{\Phi}$ is some differentiable mapping that relates the two representations. While the proposed approach permits many possible parameterized representations, here we will focus

on the case where the SOS at each pixel in the simulation grid corresponds to one of Q possible values. In this case, $\mathbf{\Phi} \in \mathbb{R}^{N \times Q}$ is a binary matrix defined as

$$[\mathbf{\Phi}]_{i,j} \equiv \begin{cases} 1 & i \in \mathcal{I}_j \\ 0 & otherwise \end{cases}, \tag{4.2}$$

where $\mathcal{I}_j$ denote the set of indices of pixels that correspond to the j-th parameterized SOS value, and $[\mathbf{A}]_{i,j}$ denotes the (i,j) element of the matrix $\mathbf{A}$.

4.3 Joint reconstruction

Given some choice for $\mathbf{\Phi}$, the parameterized JR problem is given by

$$\widehat{\mathbf{p}_0}, \widehat{\mathbf{c}_p} = \underset{\mathbf{p}_0 \geq 0, \mathbf{c}_p}{\arg\min} \, F_{PA}\left(\mathbf{p}_0, \mathbf{\Phi}\mathbf{c}_p\right) + \lambda R\left(\mathbf{p}_0\right). \tag{4.3}$$

where F_{PA} is the data fidelity term given in Eqn. 2.32, $\mathcal{R}$ is a regularization function for the initial pressure distribution, and λ is the corresponding regularization parameter value. Use of the adjoint state method to calculate the gradient of F_{PA} with respect to $\mathbf{p}_0$ and $\mathbf{c}$ is detailed in Appendix B. With that, the gradient of the data fidelity term with respect to the parameterized SOS distribution can be related to the gradient with respect to the SOS over the whole domain via the chain rule. In particular, when $\mathbf{\Phi}$ corresponds to a real-valued linear mapping, the gradient with respect to the parameterized SOS distribution is given by

$$\nabla_{\mathbf{c}_p} F_{PA}\left(\mathbf{p}_0, \mathbf{\Phi}\mathbf{c}_p\right) = \mathbf{\Phi}^T \nabla_{\mathbf{c}} F_{PA}\left(\mathbf{p}_0, \mathbf{\Phi}\mathbf{c}_p\right). \tag{4.4}$$

For the choice of $\mathbf{\Phi}$ given by Eqn.4.2, this expression reduces to

$$\left[\nabla_{\mathbf{c}_p} F_{PA}\left(\mathbf{p}_0, \mathbf{\Phi}\mathbf{c}_p\right)\right]_j = \sum_{i \in \mathcal{I}_j} \left[\nabla_{\mathbf{c}} F_{PA}\left(\mathbf{p}_0, \mathbf{\Phi}\mathbf{c}_p\right)\right]_i . \tag{4.5}$$

Since the gradients with respect to the initial pressure and SOS distributions are calculated via the adjoint state method as described in Appendices A and B, only two wave solver runs are needed to compute both gradients.

Based on Eqn. 4.5, the magnitude of j-th component of $\nabla_{\mathbf{c}_p} F_{PA}$ will depend on the cardinality of the set $\mathcal{I}_j$ or equivalently the number of pixels that correspond to the j-th component of $\mathbf{c}_p$. As a result, the gradient may be poorly scaled, leading to slow convergence. To combat this, a diagonal, positive definite scaling matrix is introduced, defined as

$$\left[\mathbf{B}\right]_j = \frac{1}{|\mathcal{I}_j|}. \tag{4.6}$$

This matrix can be employed to generate a new search direction $\mathbf{B}\nabla_{\mathbf{c}_p} F_{PA}$ that represents the average gradient over the pixels for each component of the parameterized SOS distribution. This new search direction is always a descent direction as $\mathbf{B}$ is positive-definite by construction. Other choices for $\mathbf{B}$ could result in even faster convergence; however, the choice given by Eqn. 4.6 requires little additional computational cost and ensures that the magnitude of each gradient component does not depend strongly on the number of pixels corresponding to each parameter.

With this, Eqn. 4.3 may be solved by the weighted proximal gradient descent method summarized in Algorithm 4. This algorithm can be split into two major parts, one in which the initial pressure distribution is updated via a proximal gradient descent step and one in which the parameterized SOS distribution is updated via a weighted gradient descent step.

This separation is possible since the constraint and non-smooth regularization function $\mathcal{R}$ only involve $\mathbf{p}_0$ and since the the diagonal elements of $\mathbf{B}$ differ from one only for $\mathbf{c}_p$. For brevity, the following notation is employed to describe the algorithm,

$$F_{PA}^{(k)} \equiv F_{PA}\left(\mathbf{p}_0^{(k)},\ \mathbf{c}_p^{(k)}\right) \tag{4.7a}$$

$$\mathbf{G}_{PA,p}^{(k)} \equiv \nabla_{\mathbf{p}_0} F_{PA}\left(\mathbf{p}_0^{(k)}, \mathbf{c}_p^{(k)}\right) \tag{4.7b}$$

$$\mathbf{G}_{PA,c}^{(k)} \equiv \nabla_{\mathbf{c}_p} F_{PA}\left(\mathbf{p}_0^{(k)}, \mathbf{c}_p^{(k)}\right), \tag{4.7c}$$

where the superscript k refers to the k-th iteration of the algorithm.

Algorithm 4 Parameterized joint reconstruction method

Input: $\mathbf{c}_p^{(0)}, \mathbf{p}_0^{(0)}, \lambda$
Output: $\widehat{\mathbf{c}_p}, \widehat{\mathbf{p}_0}$

$k \leftarrow 0$ {k is the algorithm iteration number.}
while stopping criterion is not satisfied **do**
  Calculate $\mathbf{G}_{PA,p}^{(k)}$ and $\mathbf{G}_{PA,c}^{(k)}$
  Choose α_k^p and α_k^c via Algorithm 5
  $\mathbf{p}_0^{(k+1)} \leftarrow \text{prox}_{\alpha_k^p \lambda R}\left(\mathbf{p}_0^{(k)} - \alpha_k^p \mathbf{G}_{PA,p}^{(k)}\right)$
  $\mathbf{c}_p^{(k+1)} \leftarrow \mathbf{c}_p^{(k)} - \alpha_k^c \mathbf{B}\mathbf{G}_{PA,c}^{(k)}$
  $k \leftarrow k + 1$
end while
$\widehat{\mathbf{c}_p} \leftarrow \mathbf{c}_p^{(k)}$
$\widehat{\mathbf{p}_0} \leftarrow \mathbf{p}_0^{(k)}$

To choose the step sizes for updating the initial pressure and parameterized SOS distributions, a two-part backtracking line search procedure, detailed in Algorithm 5, is employed. First, a step size is chosen for updating the initial pressure distribution, following the line search method outlined in [12,84] for proximal methods. During this step, the parameterized SOS distribution is kept fixed at its current value. Then, a step size for updating the parameterized SOS distribution is chosen via a conventional backtracking line search. During this

step, the cost function is evaluated at the updated initial pressure value by use of the step size found during the earlier line search. In some cases under this procedure, the previously calculated gradient for the parameterized SOS distribution may no longer represent a descent direction at the updated initial pressure distribution value. This could occur if the gradient with respect to the parameterized SOS distribution changes rapidly for small variations in the initial pressure distribution. To address this case, the update to the parameterized SOS distribution is skipped if a suitable step size cannot be found in l_{max} line search steps. These steps ensure that the chosen step sizes lead to a decrease in the value of the cost function, but may lead the parameterized SOS distribution to be updated less frequently. This issue could be avoided by recalculating the gradient with respect to the parameterized SOS distribution at the updated initial pressure value or by performing a grid search over the two step sizes. Whether the additional computational cost for either of these options is justified will depend on how sensitive the gradient with respect to the parameterized SOS distribution is to changes in the initial pressure distribution, which may be problem-dependent.

4.4 Description of computer-simulation studies

The proposed approach was validated through the use of two-dimensional computer-simulation studies. While most studies were performed with a piece-wise constant numerical mouse phantom, an additional study was performed with a vessel-like numerical phantom in order to investigate the impact of high frequency content in the measured pressure on the inverse problem.

Algorithm 5 Two-parameter backtracking line search

Input: $\alpha_0^p,\ \alpha_0^c,\ F_{PA}^{(k)}, \mathbf{G}_{PA,p}^{(k)},\ \mathbf{G}_{PA,c}^{(k)}, l_{max}$

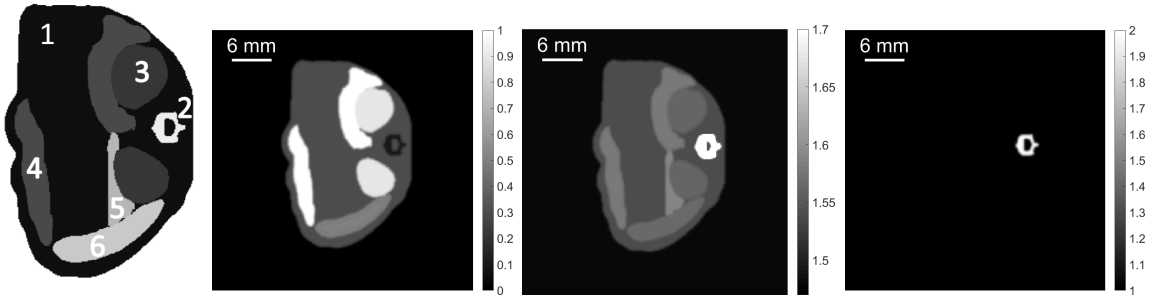
Output: $\widehat{\alpha_k^p}, \widehat{\alpha_k^c}$

$\widehat{\alpha_k^p} \leftarrow \alpha_0^p$
found $\leftarrow$ **false**
while not found **do**
  $\widetilde{\mathbf{p}_0} \leftarrow \text{prox}_{\widehat{\alpha_k^p}\lambda R}\left(\mathbf{p}_0^{(k)} - \widehat{\alpha_k^p}\mathbf{G}_{PA,p}^{(k)}\right)$
  if $F_{PA}\left(\widetilde{\mathbf{p}_0}, \mathbf{c}_p^{(k)}\right) > F_{PA}^{(k)} + \left\langle \mathbf{G}_{PA,p}^{(k)}, \widetilde{\mathbf{p}_0} - \mathbf{p}_0^{(k)} \right\rangle + \frac{1}{2\widehat{\alpha_k^p}} \|\widetilde{\mathbf{p}_0} - \mathbf{p}_0^{(k)}\|_2^2$ **then**
    $\widehat{\alpha_k^p} \leftarrow \widehat{\alpha_k^p}/2$
  else
    found $\leftarrow$ **true**
  end if
end while
$l \leftarrow 0$ {l is the line search step number.}
$\widehat{\alpha_k^c} \leftarrow \alpha_0^c$
found $\leftarrow$ **false**
while not found **and** $l < l_{max}$ **do**
  $\widetilde{\mathbf{c}_p} \leftarrow \mathbf{c}_p^{(k)} - \widehat{\alpha_k^c}\mathbf{B}\mathbf{G}_{PA,c}^{(k)}$
  if $F_{PA}\left(\widetilde{\mathbf{p}_0}, \mathbf{c}_p^{(k)}\right) < F_{PA}\left(\widetilde{\mathbf{p}_0}, \widetilde{\mathbf{c}_p}\right)$ **then**
    $\widehat{\alpha_k^c} \leftarrow \widehat{\alpha_k^c}/2$
  else
    found $\leftarrow$ **true**
  end if
  $l \leftarrow l + 1$
end while
if l is l_{max} **then**
  $\widehat{\alpha_k^c} \leftarrow 0$
end if

4.4.1 Methods

Numerical mouse phantom

The chosen numerical phantoms are based on tissue type maps generated from segmented μCT images of a mouse produced by the DigiMouse project [26]. The phantoms for the initial pressure and SOS distributions were produced by assigning each segmented region a constant initial pressure value and a constant SOS value, respectively, based on the tissue type (see Fig. 4.1). The values for the initial pressure were chosen based on the relative concentration of blood in each tissue type, as hemoglobin is one of the strongest endogenous optical absorbers in the body over the visible and near-infrared wavelength ranges [49]. The impact of optical fluence variations were not considered in this work. For most studies, unless otherwise noted, the mass density was assumed to be homogeneous. The values for the initial pressure and the SOS for each tissue type are summarized in Table 4.1.



(a) (b) (c) (d)

Figure 4.1: (a) A schematic of the segmented tissue types within the mouse. The labels for each numbered tissue type are given in Table 4.1. Phantoms for (b) the normalized initial pressure distribution, given in arbitrary units, (c) the SOS distribution, given in units of mm/μs, and (d) the mass density distribution, given in units of mg/mm^3. In order to better visualize the soft tissue variations for the SOS distribution, the grayscale window for this phantom was set to [1.47, 1.7] mm/μs resulting in saturation of the bone, which has a SOS value of 3.198 mm/μs.

Index	Region	Initial pressure	Speed of sound [mm/μs]	Mass density [mg/mm^3]
0	Water	0.0	1.480	1.00
1	Bulk tissue	0.3	1.540	1.00
2	Bone	0.1	3.198	1.99
3	Kidney	0.9	1.560	1.00
4	Liver	1.0	1.578	1.00
5	Pancreas	0.3	1.591	1.00
6	Spleen	0.5	1.567	1.00

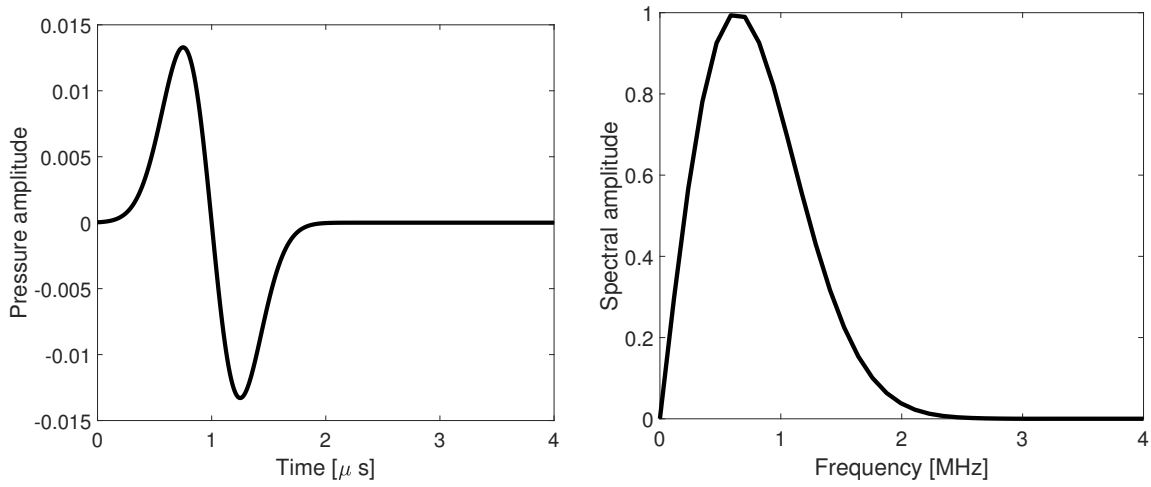
Table 4.1: Parameter values for the initial pressure, SOS, and mass density distributions for each tissue type. The normalized initial pressure values are roughly based on the relative concentration of blood for each tissue type [33, 53, 68, 102].

Imaging system

The imaging system contained 512 ultrasonic transducers evenly distributed about a ring of radius 50 mm. The transducers were treated as point-like detectors with the electro-acoustic impulse response (EIR) given in Fig. 4.2. The ultrasonic transducers were placed on the simulation grid by use of nearest neighbor interpolation.

Simulation of pressure data

In order to avoid inverse crime [20], the pressure data were simulated with different temporal and spatial sampling rates when generating the measured data and performing image reconstruction. While generating the measured data, the simulation grid consisted of a Cartesian grid with 1536×1536 pixels and a pixel size of 0.075 mm, corresponding to a physical size of approximately 115×115 mm^2. The pressure data were recorded at a temporal sampling rate of 120 MHz for 8000 time steps. Additive Gaussian white noise with zero mean and a



(a) (b)

Figure 4.2: (a) The pressure amplitude and (b) frequency spectrum of the EIR employed in the main computer-simulation studies.

standard deviation of 2% of the maximum pressure amplitude was added to the measured data. The k-Wave toolbox was employed to simulate the measured pressure data [104].

Image reconstruction

During image reconstruction, the simulation grid consisted of 768×768 pixels with a pixel size of 0.15 mm, corresponding to a physical size of approximately 115×115 mm^2. The pressure wavefield was simulated at a temporal sampling rate of 60 MHz for 4000 time steps. During the image reconstruction step, the pressure was simulated by use of a C-based numerical wave solver, implemented using NVIDIA's CUDA framework [1]. The gradients with respect to the initial pressure and parameterized SOS distributions were estimated within a field-of-view of radius 20 mm.

4.4.2 Fixed constant speed of sound

Initial pressure distributions were reconstructed by solving Eqn. 2.31 for several fixed constant SOS values and choices of λ values. The stopping criterion was when the change in the ℓ_2-norm of the initial pressure distribution between successive iterations was less than 10^{-4}.

Most traditional image reconstruction algorithms in PACT assume a constant SOS. In practice, the value of this constant SOS is often tuned according to some criteria. Three such criteria were considered for tuning a constant SOS value: (1) the root-mean-square error (RMSE), (2) the value of the cost function given in Eqn. 2.31, and (3) the Tenenbaum sharpness [105]. Use of the Tenenbaum sharpness for tuning a constant SOS value was previously proposed by Treeby, et al. [105]. Joint reconstruction with a 1-parameter SOS model can be viewed as choosing the SOS value according to the value of the cost function. While the best criteria to employ to select a constant SOS value will depend on the task or tasks for which the image will be utilized [10], here, only a few simple measures are considered.

4.4.3 Joint reconstruction

Several different SOS parameterizations were considered, each with a different number of parameters (see Table 4.2). For this initial investigation, the boundaries of these regions were based on the true boundaries of the different tissue types shown in Fig. 4.1a. When the assumed parameterization is too simple to describe the true SOS variations within the object, model error will lead to artifacts in the reconstructed images. When the assumed parameterization is very complex, the inverse problem may be poorly conditioned and may suffer from many local minima or saddle points. By considering a collection of different

parameterizations with different number of parameters, this trade-off between model error and the conditioning of the inverse problem is investigated.

Q	Regions
1	Constant SOS
2	(1) Background, (2) Mouse
3	(1) Background, (2) Soft tissue, (3) Bone
7	(1) Background, (2) Bulk tissue, (3) Bone, (4) Kidney, (5) Liver, (6) Pancreas, (7) Spleen
55869	All pixels within field-of-view

Table 4.2: Different SOS parameterizations employed during image reconstruction.

The initial guess for initial pressure distribution was the vector of all zeros. The initial guess for the parameterized SOS distribution depended on the number of parameters and the corresponding tissue types. The initial guesses for the different tissues were 1.48 mm/μs for the background, 1.50 mm/μs for all soft tissues, and 3.00 mm/μs for the bone. For fairness of comparison, the initial guess for the 55869-parameter model was chosen such that the initial guess for SOS distribution, namely $\mathbf{\Phi c}_p^{(0)}$, was the same as for the 3-parameter and 7-parameter models, even though this high-dimensional model does not suggest any prior knowledge of the different tissue types present in the object.

The numerical wave solver can become inaccurate when there are high spatial frequencies in the SOS distribution [103]. To avoid sharp discontinuities between the different parameterized regions, the SOS distribution was smoothed with a Gaussian filter with a standard deviation of one pixel prior to running the numerical wave solver.

The stopping criterion was when the change in the ℓ_2-norm of the object between successive iterations was less than 10^{-4}. In this case, the object refers to the concatenation of the initial pressure and parameterized SOS distributions.

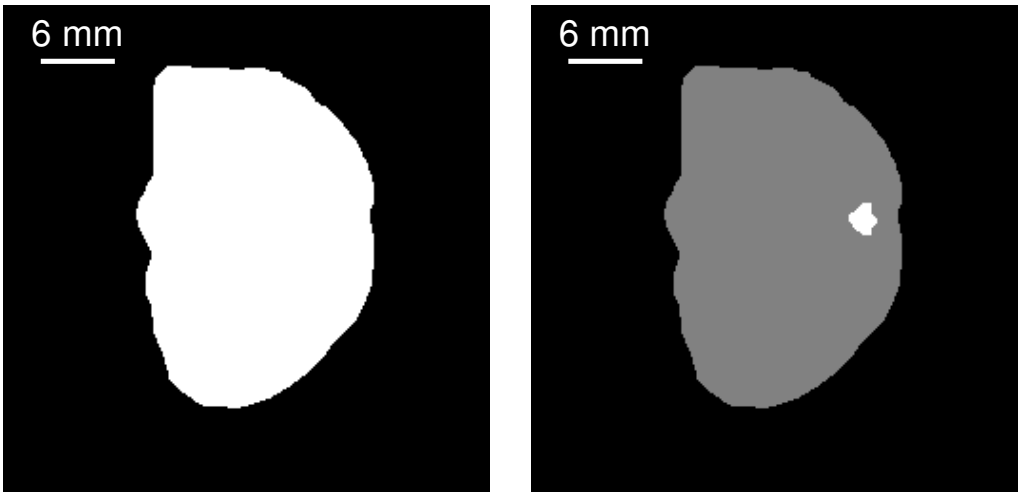
4.4.4 Mass density variations

For most studies, mass density variations are ignored during the JR process. While the derived methods can naturally be extended to include mass density variations, for simplicity, the mass density is assumed to be homogeneous. However, this assumption may be poor when bone or air voids are present in the object. To better understand the impact of ignoring mass density variations, the measured data were simulated with the heterogeneous mass density distribution shown in Fig. 4.1d. The values for mass density for each of the tissue types are summarized in Table 4.1. During image reconstruction, a homogeneous mass density distribution was assumed.

4.4.5 Imperfect parameterization

For the prior studies, it was assumed the boundaries between the different tissue types were known exactly. However, in practice, this is not the case. To better understand the impact of errors in the assumed parameterization, the boundaries of the different tissue types were estimated by segmenting a pair of reconstructed initial pressure distribution images. Two approximate parameterizations, shown in Fig. 4.3, were considered. The first was a two-parameter SOS model, obtained by segmenting the outer boundary of the mouse from the reconstructed initial pressure image obtained by JR with a one-parameter SOS model. The segmented region was determined by thresholding the initial pressure image with a cutoff of 0.1. The second model was a three-parameter SOS model. The outer boundary of the mouse was the same as for the first approximate parameterization. The outer boundary of the bone was estimated by manually segmenting a reconstructed initial pressure distribution image obtained by assuming a fixed constant SOS value of 1.51 mm/μs. It was observed

that the boundary of the bone could be more readily determined for this SOS value, which was higher than that obtained by JR.



(a) (b)

Figure 4.3: Estimated segmented regions for (a) a two-parameter SOS model and (b) a three-parameter SOS model.

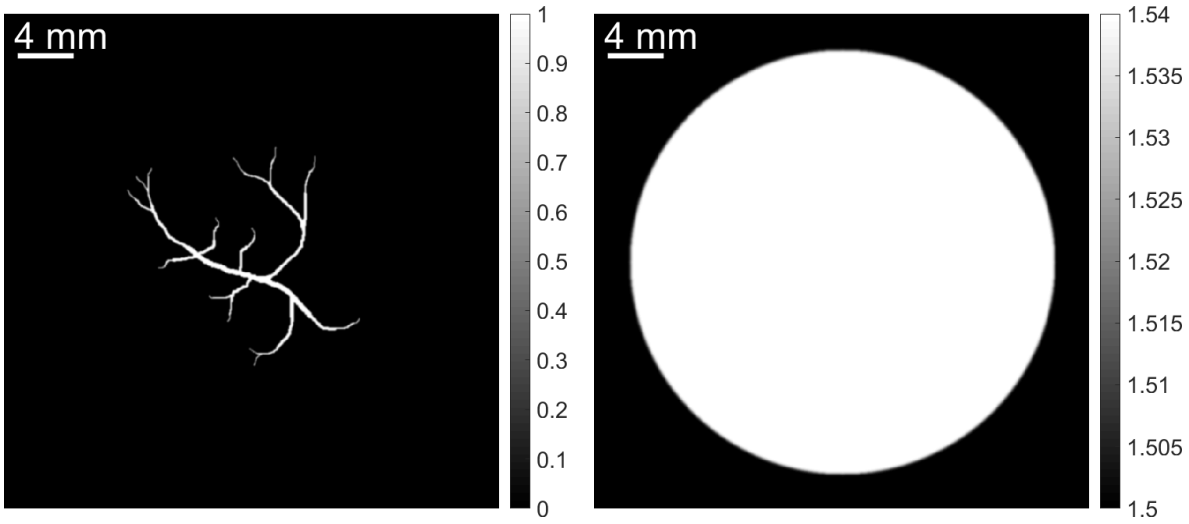
4.4.6 High-frequency pressure data

Due to the nature of the chosen initial pressure phantom, shown in Fig. 4.1b, the measured data contained predominately low-frequency content. Accurate estimation of the SOS distribution, and in turn accurate JR, can be challenging from high-frequency data due to the phenomenon of cycle skipping [14, 108]. Cycle skipping occurs when the cumulative phase error for a given frequency component exceeds π (or, equivalently, when the traveltime error exceeds half of the period). In this case, the optimization method may align a peak (or trough) in the simulated data with a neighboring peak in the measured data rather than with the true corresponding peak.

Several methods have been employed to address cycle skipping in USCT and geophysics. When a frequency-domain wave solver is utilized, a common strategy is to use a frequency hopping method in which progressively higher frequencies are used to estimate the SOS distribution [97]. For time-domain wave solvers, a similar result can be achieved by low-pass filtering the measured data or by use of multi-scale methods [14, 16].

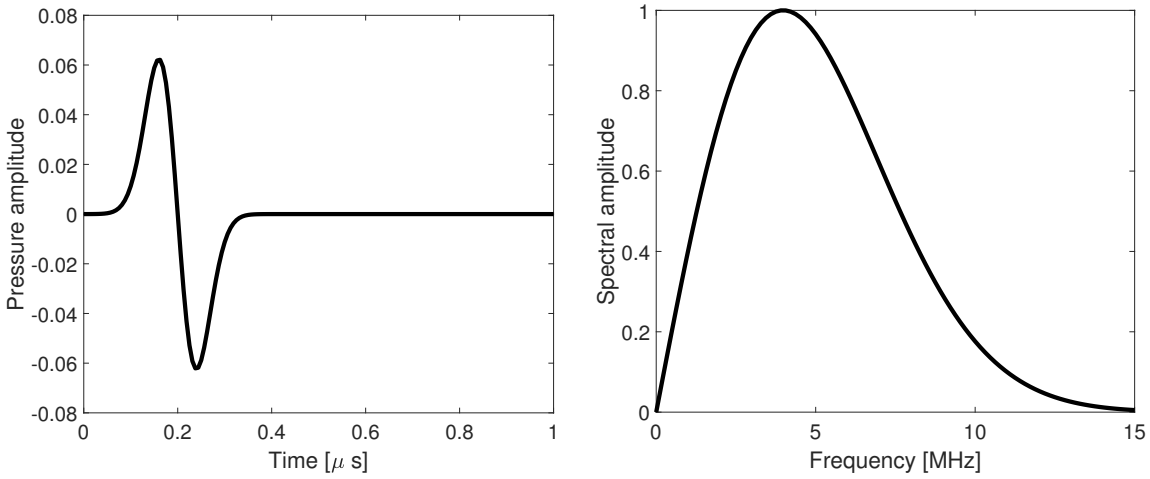
To investigate the utility of these approaches for parameterized JR, the topology of the data fidelity term was determined for measured data that were filtered with several low-pass filters with different cutoff frequencies.

To obtain high-frequency measured data, a new vessel-like numerical phantom, which contains much finer structures, was utilized. This phantom was included with the k-Wave toolbox [104]. The SOS phantom consisted of a circle with a radius of 15 mm and SOSs of 1.50 mm/μs for the background and 1.54 mm/μs for the tissue-mimicking circle. Both phantoms are shown in Fig. 4.4. In order to better capture the high-frequency content generated by the new phantom, a new EIR, shown in Fig. 4.5, with a higher central frequency was also needed. The measured pressure was simulated on a grid with 2048×2048 pixels and a pixel size of 0.05 mm, corresponding to a physical size of approximately 102 × 102 mm^2. The pressure data were recorded at a sampling rate of 160 MHz for 8000 time points. Additive Gaussian white noise with zero mean and a standard deviation of 0.5% of the maximum of the measured pressure was added to the data. The data were then filtered with several Hann low-pass filters with cutoff frequencies of 2 MHz, 4 MHz, and 6 MHz. The filtered data were downsampled to a sampling rate of 80 MHz. The value of data fidelity term was evaluated by employing a simulation grid of 1024×1024, a pixel size of 0.1 mm, a sampling rate of 80 MHz, and 4000 time points. The data fidelity term was evaluated for a range of tissue SOS values and a fixed background SOS equal to its true value. Separately, the data fidelity



(a) (b)

Figure 4.4: Phantoms for (a) the initial pressure distribution, given in arbitrary units, and (b) the SOS distribution, given in units of mm/μs, for the high-frequency computer-simulation studies.



(a) (b)

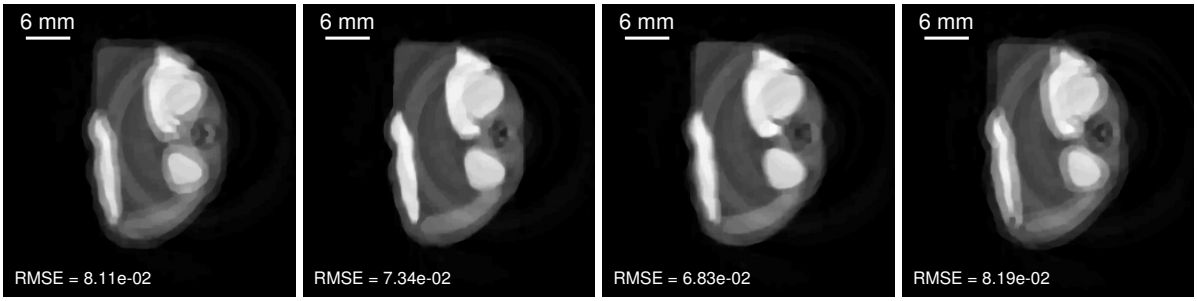
Figure 4.5: (a) The pressure amplitude and (b) frequency spectrum of the EIR employed in the high-frequency computer-simulation studies.

term was evaluated for a range of background SOS values for a fixed tissue SOS equal to its true value. In both cases, the initial pressure distribution was set to its true value.

4.5 Results of computer-simulation studies

4.5.1 Fixed constant speed of sound

The initial pressure distributions were reconstructed for several fixed constant SOS values and $\lambda = 10^{-3}$ (see Fig. 4.6). The RMSEs of each image are shown in the lower left-hand corner. The value of λ was tuned in order to minimize the RMSEs for each SOS.



(a) (b) (c) (d)

Figure 4.6: Reconstructed initial pressure distributions for $\lambda = 10^{-3}$ and fixed constant SOS values of (a) 1.480 mm/μs, (b) 1.490 mm/μs, (c) 1.500 mm/μs, and (d) 1.510 mm/μs. The images are shown in a grayscale window of [0.0, 1.1].

Three criteria for selecting the optimal constant SOS are compared: (1) the Tenenbaum sharpness [105], (2) the value of the cost function given in Eqn. 4.3, and (3) the RMSE. The values of these metrics for several constant SOS values can be found in Table 4.3. For consistency, all results are reported for the $\lambda = 10^{-3}$ case, even though the RMSE is lower for a lower value of λ for a SOS of 1.495 mm/μs.

Metric	1.485	1.49	1.495	1.50	1.505	1.51
Sharpness $\times 10^3$	2.55	3.65	**4.00**	3.04	2.65	2.60
Cost	58.9	32.7	**22.2**	33.0	55.7	78.3
RMSE $\times 10^{-2}$	7.42	7.34	7.17	**6.83**	7.15	8.19

Table 4.3: Values of different metrics that could be employed for selecting a constant SOS value. The optimal value according to each metric is shown in bold.

The sharpness and cost metrics both suggest that a constant SOS value of 1.495 mm/μs is the optimal value, while the RMSE suggests a similar, though slightly higher, value of 1.500 mm/μs. In practice, it is not possible to use the RMSE to tune the SOS as the true initial pressure distribution is unknown. Both the sharpness and cost metrics may serve as adequate proxies or alternatives. In some cases, performing JR, in which a single image is reconstructed, may be quicker than reconstructing a series of images for different SOS values and evaluating them according to some metric. It may also be possible to design the cost function such that it tracts the most relevant image quality measure.

4.5.2 Joint reconstruction

Initial pressure and parameterized SOS distributions were jointly reconstructed for the five parameterizations given in Table 4.2. The reconstructed initial pressure distributions are shown in Fig. 4.7. For consistency with the fixed constant SOS results, the value of λ was tuned to minimize the RMSE of the initial pressure distribution. The accuracy of the reconstructed initial pressure distributions is greatly improved when the chosen SOS parameterization includes compensation for the bone ($Q \geq 3$), which is the strongest source of acoustic heterogeneity in the phantom. This is can be seen in both the apparent visual quality of the images and the RMSEs. Additionally accounting for SOS variations within

the soft tissue ($Q = 7$) leads to a further minor improvement in the RMSE. The RMSE of reconstructed initial pressure distribution when the SOS parameterization allows the SOS values to vary independently for each pixel ($Q = 55869$) is worse than either the $Q = 3$ or $Q = 7$ cases. However, there is little obvious visual difference in the reconstructed initial pressure distributions for the three cases where $Q \geq 3$.

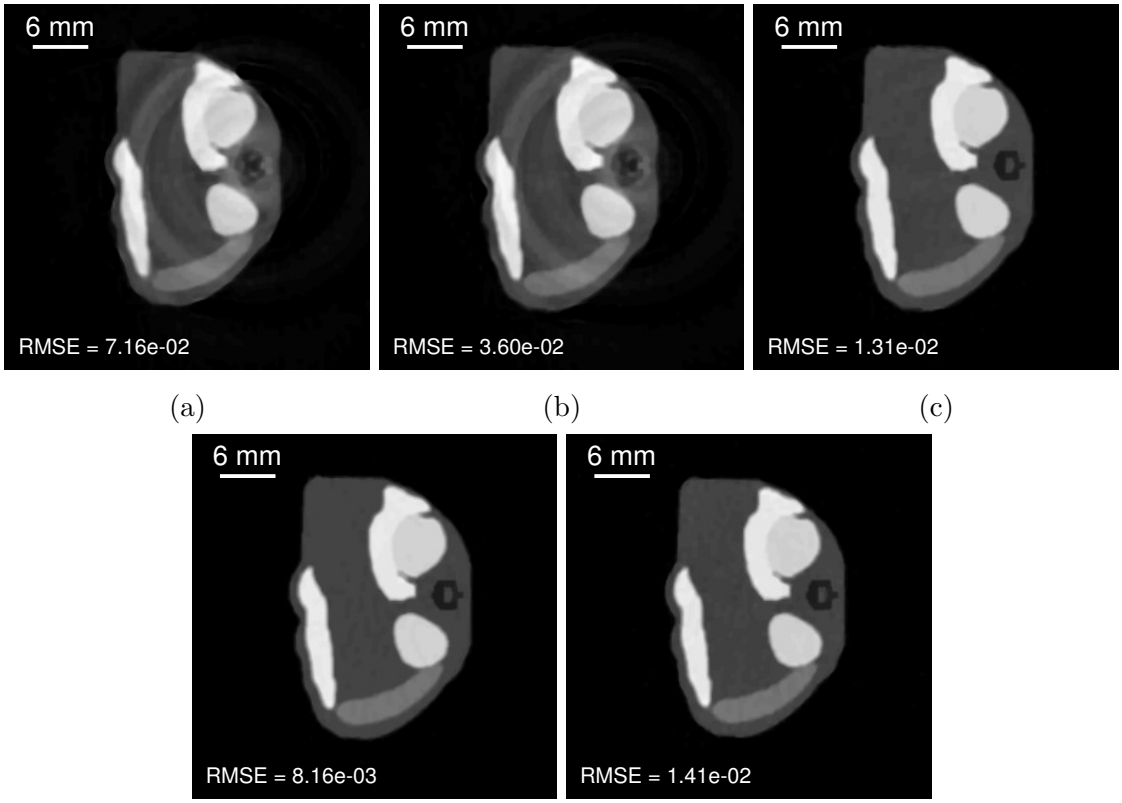


Figure 4.7: Reconstructed initial pressure distributions for parameterized JR for $\lambda = 10^{-4}$ with (a) 1 parameter, (b) 2 parameters, (c) 3 parameters, (d) 7 parameters, and (e) 55869 parameters. The images are shown in a grayscale window of [0.0, 1.1].

The corresponding reconstructed SOS distributions are shown in Fig. 4.8. The equivalent pixel-wise representations of the reconstructed parameterized SOS distributions are presented for ease of visualization and comparison with the true SOS phantom. The estimated

SOS values for $Q \leq 7$ are also summarized in Table 4.4. For the $Q = 7$ case, the estimated SOS distribution closely matches the true SOS distribution. For $Q < 7$, the true SOS distribution cannot be recovered due to the choice of the SOS parameterization. In these cases, effective SOS values are estimated for certain regions. For example, for the $Q = 2$ and $Q = 3$ cases, a single effective SOS value is estimated for the soft tissue of the mouse. This estimated SOS falls between the lower SOS value of the bulk tissue and the higher SOS values of the organs (1.555 mm/μs for $Q = 2$ and $Q = 3$). Similarly, for the $Q = 1$ case, a single effective SOS is estimated for the entire simulation grid. The value of this effective SOS is between the background SOS and the bulk soft tissue SOS of the mouse (1.495 mm/μs). When each pixel is allowed to independently vary, the estimated SOS distribution contains strong artifacts.

	Water	Bulk tissue	Bone	Kidney	Liver	Pancreas	Spleen
True	1.480	1.540	3.198	1.560	1.578	1.591	1.567
$Q = 1$	1.495	–	–	–	–	–	–
$Q = 2$	1.481	1.555	–	–	–	–	–
$Q = 3$	1.480	1.555	3.171	–	–	–	–
$Q = 7$	1.480	1.540	3.258	1.558	1.578	1.588	1.566

Table 4.4: Estimated SOS values for different parameterized SOS models. For models with $Q < 7$, the estimated SOS values represent effective SOSs across several different tissue types.

4.5.3 Mass density variations

The previous results were reconstructed from measured data that were generated assuming a constant mass density. However, the mass density of bone is very different from the mass density of soft tissue. In principle, the proposed JR approach could be employed to jointly estimate the mass density distribution along with the initial pressure and SOS distributions.

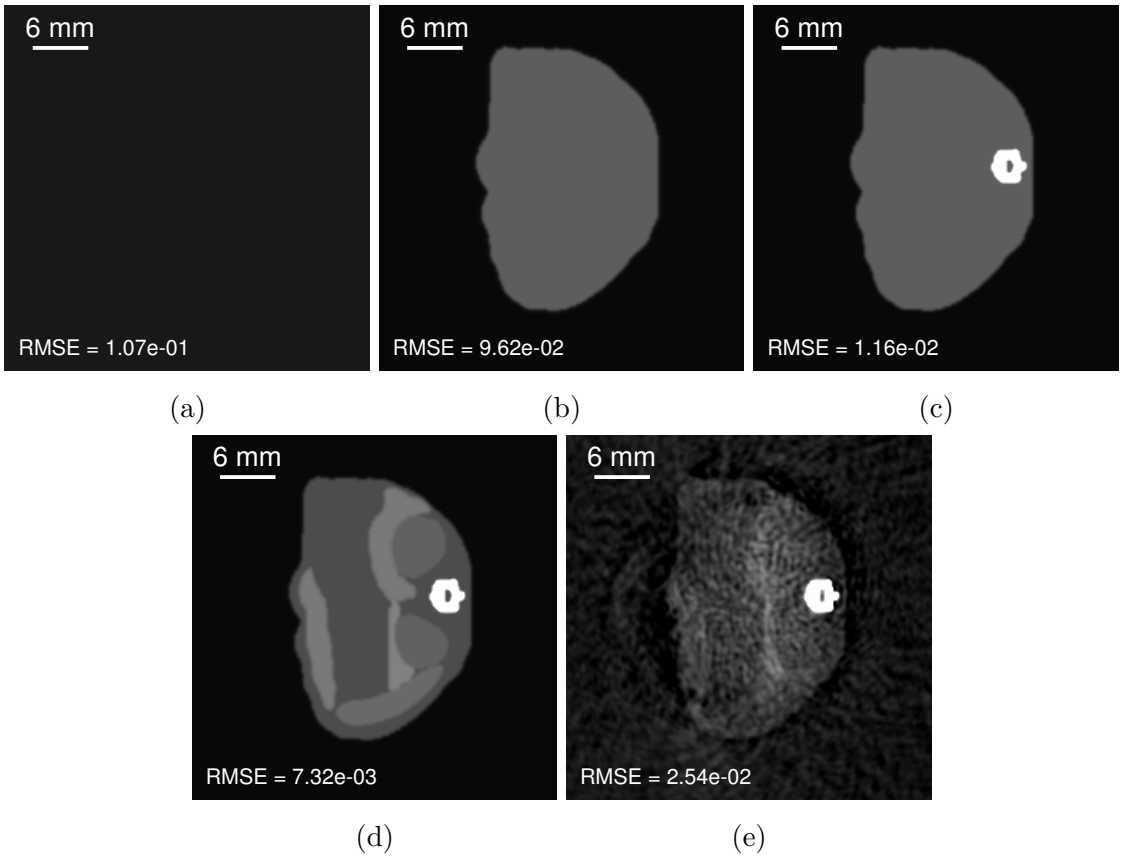
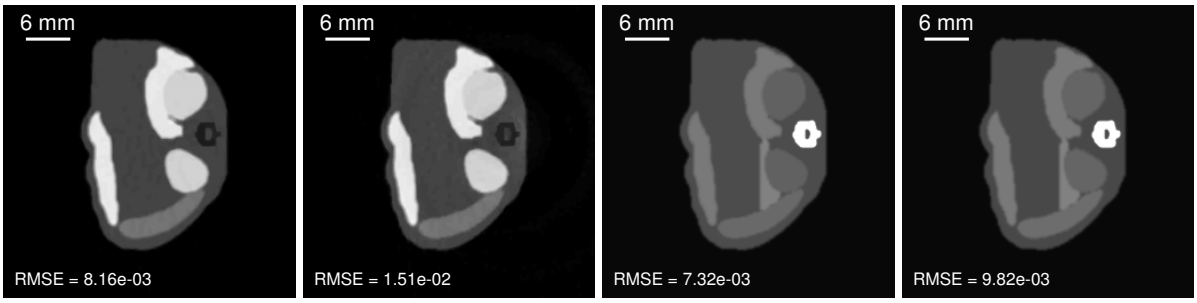


Figure 4.8: Reconstructed SOS distributions for parameterized JR for $\lambda = 10^{-4}$ with (a) 1 parameter, (b) 2 parameters, (c) 3 parameters, and (d) 7 parameters, and (e) 55869 parameters. The images are shown in a grayscale window of [1.47, 1.70] mm/μs.

However, this could result in an inverse problem that is more ill-conditioned. In order to evaluate the impact of ignoring mass density variations, the measured data were generated for the heterogeneous mass density distribution shown in Fig. 4.1d. Then, parameterized JR was performed assuming a constant mass density. The reconstructed initial pressure and SOS distributions are shown in Fig. 4.9. The initial pressure and SOS distributions reconstructed from measured data that did not include mass density variations are also shown as references. Ignoring mass density variations did have a small impact on the accuracy of the reconstructed



(a) (b) (c) (d)

Figure 4.9: Initial pressure distributions reconstructed by parameterized JR for $\lambda = 10^{-4}$ with 7 parameters from measured data that (a) ignored mass density variations and (b) included mass density variations. SOS distributions reconstructed by parameterized JR with 7 parameters from measured data that (c) ignored mass density variations and (d) included mass density variations. Mass density variations were not included in the reconstruction process. The initial pressure images are shown in a grayscale window of [0.0, 1.1]. The SOS images are shown in a grayscale window of [1.47, 1.70] mm/μs.

initial pressure and SOS distributions. In particular, the SOS value of the bone was overestimated (3.340 vs. 3.198 mm/μs). Modeling the mass density variations when generating the measured data increased the percentage of acoustic energy reflected at the bone soft tissue interface. Increasing the SOS value of bone during the reconstruction, in the absence of mass density variations, also results in more acoustic energy being reflected. This could explain the overestimation of the bone SOS.

4.5.4 Approximate segmentations

Figure 4.10 shows the reconstructed initial pressure and SOS distributions obtained by parameterized JR for the approximate tissue type segmentations shown in Fig. 4.3. As references, the reconstructed initial pressure and SOS distributions obtained by use of the true tissue type boundaries are also shown. For the $Q = 2$ case, the approximate segmentation is

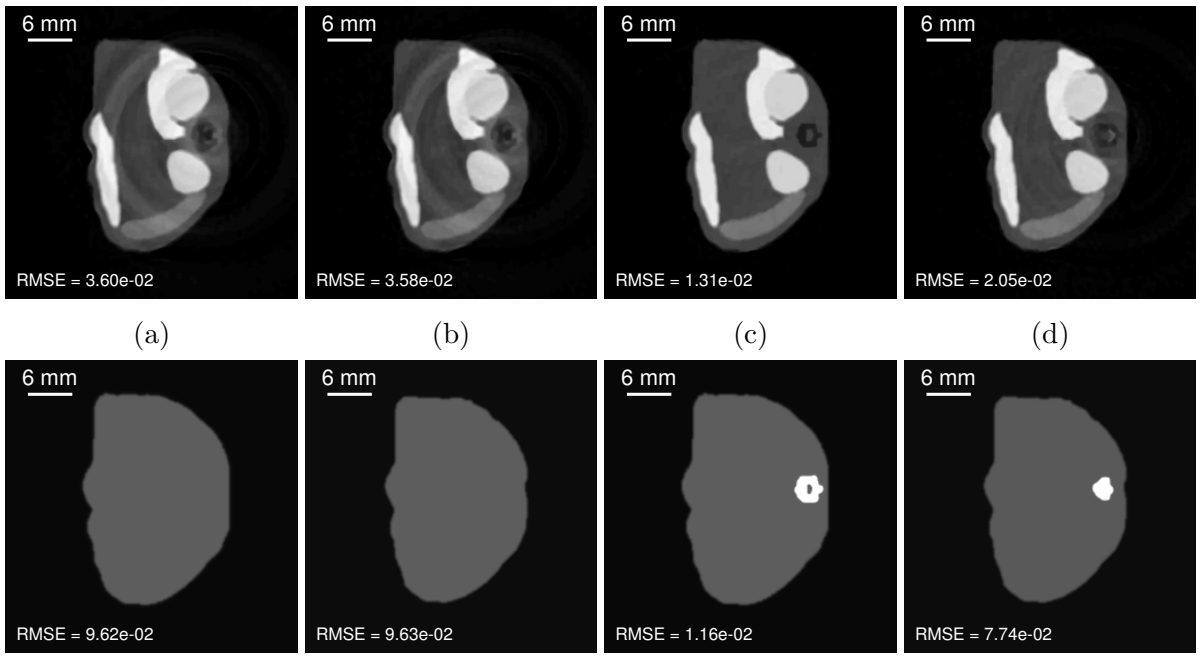


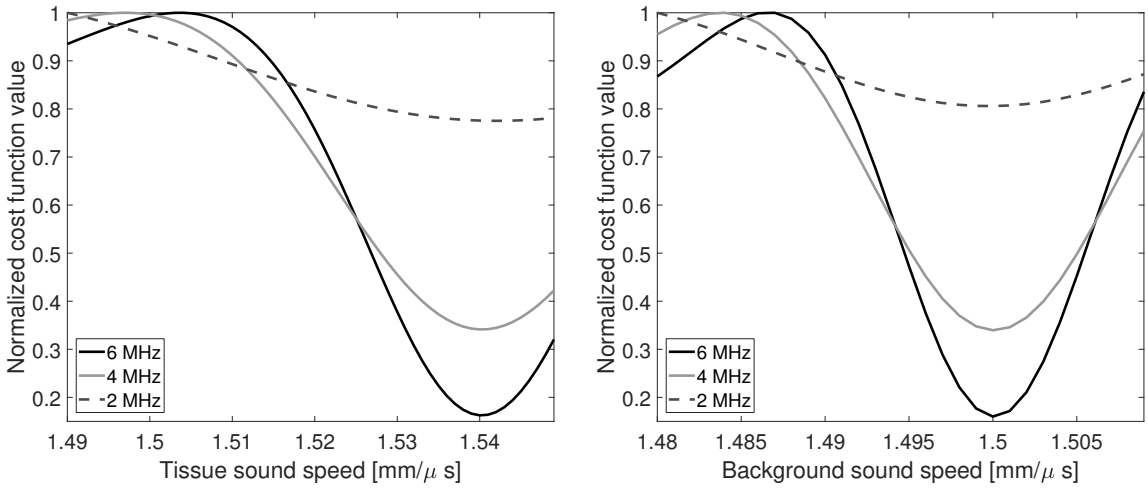
Figure 4.10: Reconstructed initial pressure distributions obtained by parameterized JR for $\lambda = 10^{-4}$ with (a) 2 parameters with perfect segmentation, (b) 2 parameters with imperfect segmentation, (c) 3 parameters with perfect segmentation, and (d) 3 parameters with imperfect segmentation. Reconstructed SOS distributions obtained by parameterized JR with (e) 2 parameters with perfect segmentation, (f) 2 parameters with imperfect segmentation, (g) 3 parameters with perfect segmentation, and (h) 3 parameters with imperfect segmentation. The initial pressure images are shown in a grayscale window of [0.0, 1.1]. The SOS images are shown in a grayscale window of [1.47, 1.70] mm/μs.

very close to the true segmentation, as the outer boundary of the mouse can be estimated relatively accurately. As a result, there is close agreement between the reconstructed images

for the exact and approximate segmentations. For the $Q = 3$ case, the outer boundary of bone could not be estimated very accurately. As a result, the differences between the reconstructed images for the exact and approximate segmentations are large. Still, the reconstructed initial pressure and SOS images for the $Q = 3$ with approximate segmentations both have lower RMSEs than the corresponding images for the $Q = 2$ case with exact parameterizations. This suggests that the parameterized JR method is at least somewhat robust to errors in the assumed parameterization.

4.5.5 Impact of frequency content

The values of the data fidelity term, evaluated at the true initial pressure distribution and the true background SOS, for different values of the tissue SOS are shown in Fig. 4.11a. The values of the cost function are normalized in order to allow comparison for measured data subjected to low-pass filters with different cutoff frequencies. The width of the basin of attraction about the true tissue SOS of 1.54 mm/μs grows wider as the cutoff frequency for the low-pass filter is reduced. This helps demonstrate how low-pass filtering the measured data can help avoid local minima. As an example, consider the case where the initial guess for the tissue SOS was equal to the background value of 1.5 mm/μs. For the data with a cutoff frequency of 6 MHz, the estimated tissue SOS value would become smaller instead of approaching its true value. This is not true for the data with lower cutoff frequencies. However, there is a downside to filtering the data too aggressively. As the cutoff frequency is decreased, the basin of attraction also becomes shallower, potentially resulting in slower convergence. This observation is consistent with prior work that recommends maximizing bandwidth while avoiding cycle skipping [108].



(a) (b)

Figure 4.11: Normalized cost function values for (a) different tissue SOSs and (b) different background SOSs after applying a Hann low-pass filter to the measured data. Results are shown for several cutoff frequencies.

Similar phenomena can be observed for the case where the background SOS is swept and the initial pressure distribution and the tissue SOS are held fixed at their true values (see Fig. 4.11b). In this case, the width of the basin of attraction is even narrower as the propagation distance through the background is longer than that through the tissue for the considered phantom. As a result, phase errors due to an incorrect background SOS can accumulate over longer distances, leading to a greater sensitivity to absolute errors in the estimated SOS.

4.6 Description of experimental studies

The utility of the proposed approach was also evaluated through experimental studies.

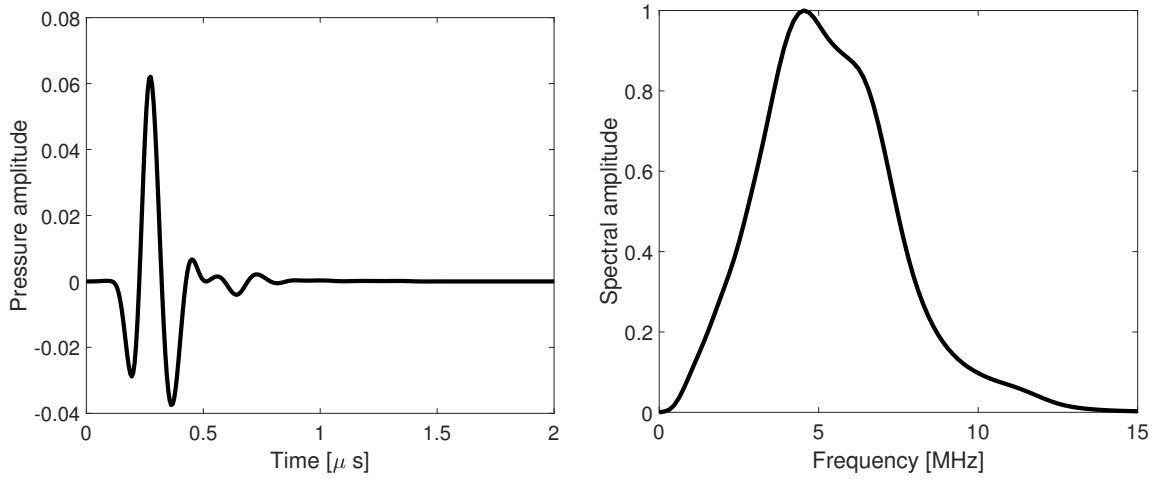
4.6.1 Methods

In vivo measurements of the trunk of an anesthetized adult nude mouse (Hsd:Athymic Nude-FoxlNU, Harlan) were previously acquired by the Optical Imaging Laboratory [65]. All experimental procedures were carried out in conformity with laboratory animal protocols approved by the Animal Studies Committee at Washington University in St. Louis.

The imaging system consisted of a ring array with a radius 50 mm and 512 evenly distributed ultrasonic transducers. The transducers were elevationally focused and had a central frequency of approximately 5 MHz. The pressure data were recorded for 2000 time steps at a sampling rate of 40 MHz. Twelve transducers were identified as having low sensitivity by manual inspection and were excluded from the image reconstruction process. A 1064 nm pulsed laser (DLS9050, Continuum) with a 5-9 ns pulse width and pulse repetition rate of 50 Hz was used to illuminate the mouse. The fluence on the surface of the mouse was approximately 18 mJ/cm^2. More detailed information on this imaging system can be found in [65].

The EIR of the ultrasonic transducers was estimated by measuring the pressure from a point-like absorber placed in the center of the ring array as described in the supplementary material of [65]. The measured pressure signals were shifted to align the peaks in order to account for small differences in the propagation distances among the transducers. Then, the frequency response of a point source was deconvolved from the measured pressure data by use of Wiener deconvolution. Finally, the responses were averaged across all transducers and the resulting estimated EIR is shown in Fig. 4.12.

The measured data were upsampled to a sampling frequency of 80 MHz by linear interpolation in order to increase the numerical stability of the wave solver [103]. The resulting data



(a) (b)

Figure 4.12: (a) The pressure amplitude and (b) frequency spectrum of the estimated EIR for the experimental system.

were then low-pass filtered with a Butterworth filter with a cutoff of 12 MHz. In addition, the measured pressure data were temporally shifted to account for any unwanted shift induced by the EIR. The size of this shift was calculated by convolving a narrow pulse with the EIR and observing the shift in the envelope of the narrow pulse.

4.6.2 Fixed speed of sound

The initial pressure distribution was estimated for several fixed constant SOS values by solving Eqn. 2.31. The simulation grid consisted of 2048×2048 pixels with a pixel size of 0.05 mm, corresponding to a physical size of approximately 102 × 102 mm^2. A temporal sampling rate of 80 MHz was employed, and the radius of the field-of-view was 40 mm. The stopping criterion was when the change in the ℓ_2-norm of the initial pressure distribution between successive iterations was less than 10^{-4}.

4.6.3 Joint reconstruction

Joint reconstruction was performed for the two-parameter SOS model shown in Fig. 4.13. This parameterized model was estimated by manually segmenting the outer boundary of the mouse from the reconstructed initial pressure distribution for a fixed constant SOS of 1.50 mm/μs.

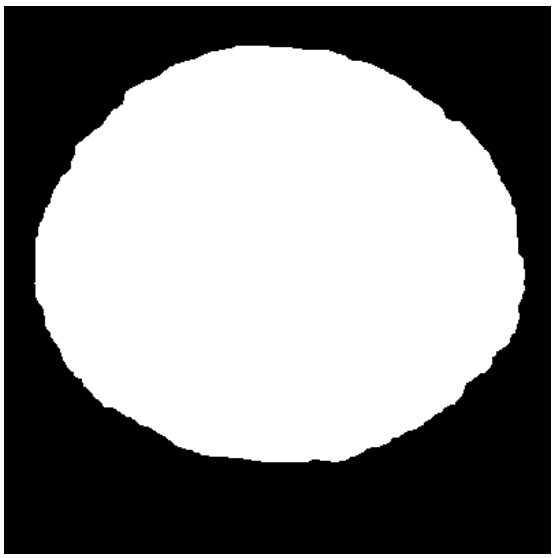


Figure 4.13: Estimated segmented regions of the mouse for a 2-parameter SOS model.

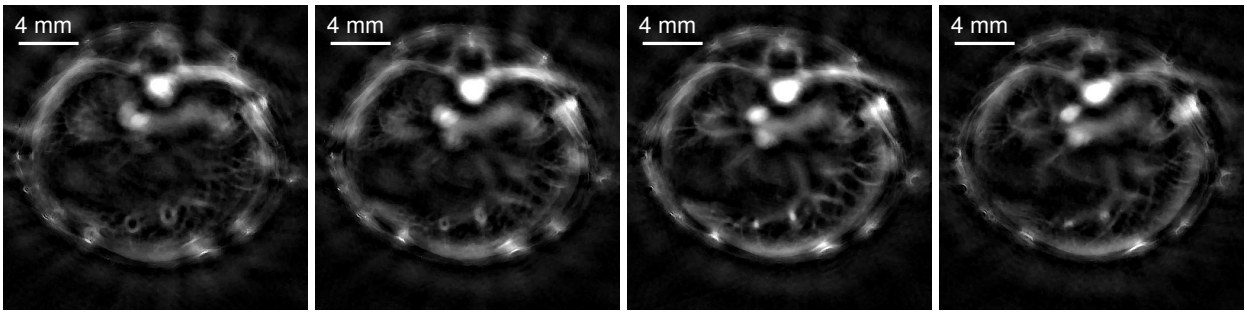
To avoid the phenomenon of cycle skipping discussed in Section 4.4.6, the JR images were estimated by a multi-stage process. First, a Hann low-pass filter with a cut-off frequency of 1 MHz was applied to the measured data. Due to the low cutoff frequency, the temporal sampling frequency was reduced from the 80 MHz to 40 MHz. The pixel size and the number of pixels in the simulation grid were also reduced to 0.1 mm and 1024×1024, respectively. The initial guess for the initial pressure distribution was the vector of all zeros, while the initial guess for the parameterized SOS distribution was 1.48 mm/μs for the background and 1.54 mm/μs for the mouse. This stage was run for 600 iterations. It was observed that additional iterations resulted in minimal changes to the estimated parameterized SOS

distribution. Second, the estimated initial pressure and SOS distributions were refined by JR with the measured data with a 12 MHz cutoff frequency. By raising the cutoff frequency, it is hoped that finer structures in the object can be recovered. The estimated initial pressure and SOS distributions obtained from the previous stage served as the initial guess for this later stage. This second stage was run for 500 iterations. Finally, the initial pressure distribution was reconstructed by use of FISTA for the fixed SOS distribution obtained during the second stage [11]. The stopping criterion was when the change in the ℓ_2-norm of the initial pressure distribution between successive iterations was less than 10^{-4}. Since the final stage employed an accelerated first-order optimization method (FISTA), faster convergence could be achieved compared with the unaccelerated proximal gradient method employed as part of the earlier stages. Use of FISTA also avoids the two-step line search method employed as part of the parameterized JR procedure.

4.7 Results of experimental studies

The initial pressure distributions were reconstructed for several fixed constant SOS values (see Fig. 4.14). Strong surface and interior vessel structures can be observed in the reconstructed images. The interior vessel structures appear most in focus for a constant SOS value of 1.500 mm/μs; however, for this SOS value, some surface vessels, particularly along the lower right side of the mouse, appear as arcs rather than points, suggesting that they are out-of-focus. As a result, there is no single constant SOS value for which all of the features of the image are in focus.

The initial pressure and SOS distributions were estimated by a multi-stage JR process. Following the first-stage, the estimated SOS values were 1.489 mm/μs for the background

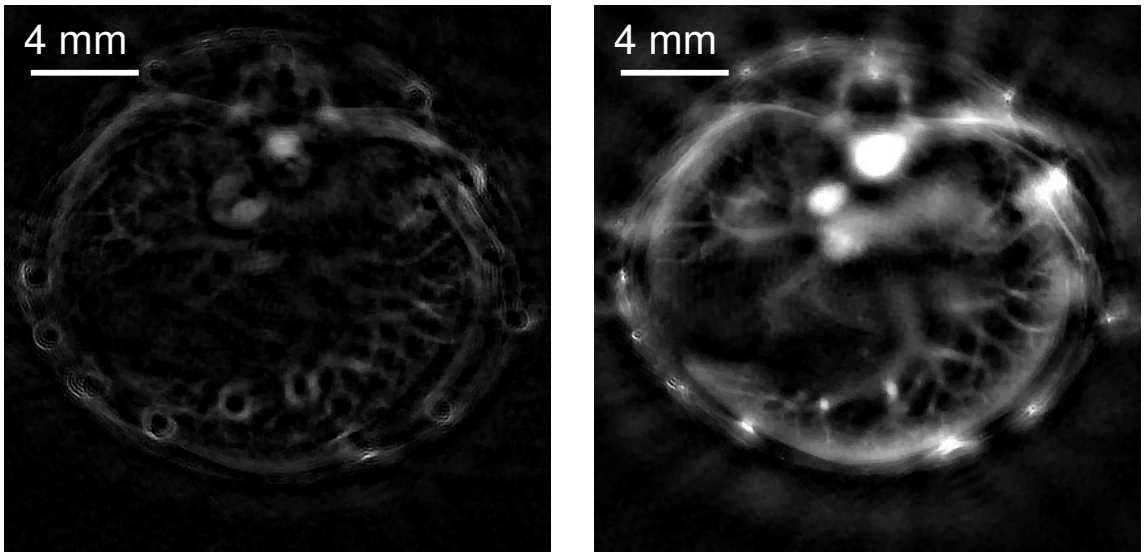


(a) (b) (c) (d)

Figure 4.14: Reconstructed initial pressure distributions for several fixed constant SOS values of (a) 1.490 mm/μs, (b) 1.495 mm/μs, (c) 1.500 mm/μs, and (d) 1.505 mm/μs. Results are shown for $\lambda = 10^{-1}$ in a grayscale window of $[0, 8000]$.

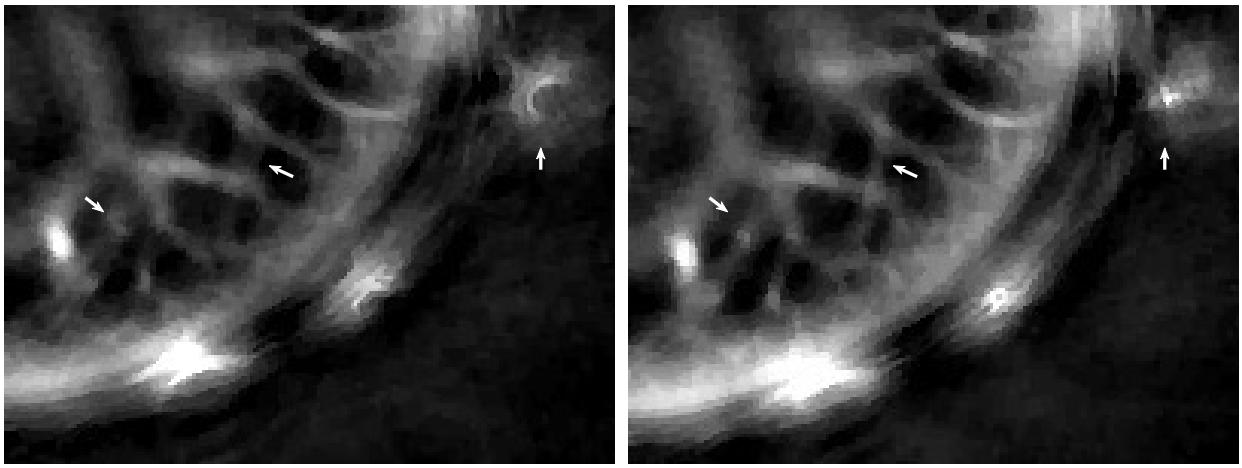
and 1.565 mm/μs for the body of the mouse. These were refined to 1.492 mm/μs for the background and 1.561 mm/μs for the mouse body during the second stage. In Fig. 4.15, the reconstructed initial pressure distribution assuming a SOS distribution equal to the initial guess for the SOS distribution and a SOS distribution equal to the final SOS distribution obtained by JR are shown. In the image obtained with the final SOS distribution, the vessels can be much more clearly visualized compared with the image obtained with the SOS corresponding to the initial guess. By employing a multi-stage reconstruction process, a focused initial pressure distribution image can be obtained even if there are moderate errors in the initial guess for the parameterized SOS distribution.

The initial pressure distribution image obtained by JR also demonstrates improvement over the image obtained with a tuned constant SOS. For example, consider the zoomed-in regions of the reconstructed initial pressure distributions shown in Fig. 4.16. Several structures appear more in focus in the JR image. In particular, the rightmost surface vessel appears as an arc in the constant SOS image and a point in the JR image. Additionally, several interior vessels are better focused in the JR image.



(a) (b)

Figure 4.15: Reconstructed initial pressure distributions for $\lambda = 10^{-1}$ assuming (a) the SOS distribution employed as the initial guess for the first stage of JR and (b) the SOS distribution obtained by parameterized JR for $Q = 2$. Results are shown in a grayscale window of $[0, 8000]$.



(a) (b)

Figure 4.16: Zoomed-in region of the reconstructed initial pressure distributions for $\lambda = 10^{-1}$ assuming (a) a tuned constant SOS of 1.500 mm/μs and (b) the SOS distribution obtained by parameterized JR for $Q = 2$. The arrows point to structures that are in focus in the JR image, but not in the tuned constant SOS image. Results are shown in a grayscale window of $[0, 6000]$.

4.8 Conclusions

Parameterized JR can produce accurate estimates of the initial pressure distribution and a low-dimensional representation of the SOS distribution from PACT measurements alone. Use of a low-dimensional parameterized form for the SOS distribution can help stabilize the JR problem. Further, this approach can result in more accurate reconstructed initial pressure distributions than assuming a constant SOS. Parameterized JR may also offer advantages compared with manually tuning a parameterized SOS model, particularly when the number of parameters in the SOS model is large.

The proposed approach is one example of the possible benefits of employing different discretizations of the object for the forward and inverse problems. By choosing the discretizations independently, the representation of the object may be individually tailored for the intended task. For example, the numerical wave solver may require a small pixel size in order to limit errors in the simulated pressure, leading to a high-dimensional representation of the SOS distribution [103]. However, a low-dimensional representation for the SOS distribution may be more beneficial for the inverse problem as the number of unknowns is reduced.

Parameterized JR does not eliminate all of the challenges associated with JR of the initial pressure and SOS distributions in a general setting. Avoiding local minima can be difficult, particularly for the case of high-frequency measured data. Additional practical methods for determining useful parameterizations for the SOS distribution in an experimental setting are needed. The cost function could be more carefully designed in order to ensure that minimizing the cost function produces images that maximize the most relevant image quality measures. The assumed acoustic model also has a number of limitations. For example, a 2-D

imaging model is assumed, the spatial impulse response of the transducers is ignored, and acoustic attenuation and mass density variations are neglected. In addition, the proposed method is somewhat computationally expensive. Use of more efficient optimization methods for solving the JR problem, such as accelerated first-order or second-order methods, may reduce this computational burden and thus increase the attractiveness of this approach.

Chapter 5

Joint reconstruction of the speed of sound and initial pressure distributions from combined PACT/USCT data

5.1 Overview

As mentioned in Chapter 4, most traditional image reconstruction methods for PACT assume that the SOS is constant throughout the medium. This assumption does not hold for many biological imaging applications [45, 88, 120]. Failure to account for variations in the SOS distribution can give rise to artifacts in the reconstructed initial pressure distribution. Several PACT image reconstruction methods, including time-reversal [44, 106, 121], iterative full-wave inversion [7, 46], and the Neumann-series or iterative time reversal approach [91, 101], have been proposed to account for known SOS variations. However, in practice, the SOS

distribution is unknown. Further, it has been shown that, in general, JR of the SOS and initial pressure distributions from PACT measurements alone is unstable [47, 100]. While use of a low-dimensional parameterization for the SOS distribution can help mitigate this instability (as seen in Chapter 4), determination of an appropriate parameterization may not always be straight-forward.

One alternative is to employ adjunct imaging data to estimate the SOS distribution when such data is available. As demonstrated in Chapter 3 and shown in several previous studies, USCT permits accurate reconstruction of the SOS distribution for breast imaging [17, 27, 39, 42, 64, 93, 113]. Further, USCT and PACT commonly employ similar detection hardware and similar imaging geometries. Consequently, combined USCT/PACT imaging systems have become an emerging area of exploration [31, 70, 115]. These integrated systems offer several advantages over independent or single-modality systems. First, they can produce automatically co-registered images with both optical and acoustic contrasts. These complementary contrasts could aid in a variety of medical imaging tasks, including breast cancer detection. Second, integration of these modalities should be comparatively simple and inexpensive due to their similar detection hardware. Third, an improved estimate for the SOS distribution, enabled by the USCT data, may allow the initial pressure distribution to be more accurately reconstructed in PACT.

In cases where combined USCT/PACT systems have been previously employed, the SOS and initial pressure distributions were reconstructed in a sequential manner. First, the SOS distribution was estimated from USCT data alone. Then, the initial pressure distribution was estimated from the PACT data alone by use of the previously estimated SOS distribution. As will be shown, this approach is not optimal. Since the photoacoustic waves propagate according to the acoustic properties of the medium, the measured PACT data

contain information on the SOS distribution. This information is ignored under the sequential approach, which estimates the SOS distribution from USCT measurements alone. Here, a JR approach that estimates the initial pressure distribution and the SOS distribution from combined PACT and USCT measurements is proposed. This synergistic approach offers dual benefits over the conventional approach: (1) it effectively utilizes the acoustic information in the PACT data, allowing the SOS distribution to be accurately estimated from few-view USCT measurements, and (2) it allows the initial pressure distribution to be accurately estimated by providing a more accurate estimate of the SOS distribution than could be obtained from PACT measurements alone.

5.2 Joint reconstruction

Instead of estimating the SOS distribution from USCT measurements alone, as described in Section 2.2.3, the SOS distribution can be estimated from combined PACT and USCT measurements in order to leverage the acoustic information present in the PACT measurements. This can be accomplished by jointly estimating the SOS and initial pressure distributions as

$$\hat{\mathbf{p}_0}, \hat{\mathbf{c}} = \underset{\mathbf{p}_0 \geq 0, \mathbf{c}}{\arg\min} \, F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right) + \beta F_{SUS}\left(\mathbf{c}\right) + \lambda_p \mathcal{R}_p\left(\mathbf{p}_0\right) + \lambda_c \mathcal{R}_c\left(\mathbf{c}\right), \tag{5.1}$$

where F_{PA} is the data fidelity term for PACT defined in Eqn. 2.32, F_{SUS} is the data fidelity term for USCT defined in Eqn. 2.41, $\beta > 0$ is a scalar that controls the relative weight of the two data fidelity terms, $\mathcal{R}_p$ is the regularization term for the initial pressure distribution, λ_p is the corresponding regularization parameter, $\mathcal{R}_c$ is the regularization term for the SOS distribution, and λ_c is the corresponding regularization parameter. Based on the approach suggested by Huang, et al. [47] for JR from PACT measurements alone, this optimization

problem is solved by use of an alternating minimization approach. As described in Algorithm 6, this approach consists of solving two subproblems in an alternating fashion. For the first subproblem, the initial pressure distribution is estimated for a fixed SOS distribution from PACT measurements alone. This problem is identical to the one previously discussed in Section 2.2.1. For the second subproblem, the SOS distribution is estimated from combined USCT and PACT measurements for a fixed initial pressure distribution as

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} F_{PA}\left(\mathbf{p}_0, \mathbf{c}\right) + \beta F_{SUS}\left(\mathbf{c}\right) + \lambda_c \mathcal{R}_c\left(\mathbf{c}\right). \tag{5.2}$$

This subproblem can be solved by SGD by employing a combination of the procedures described in Chapters 3 and 4. For this approach, the gradients with respect to the SOS distribution are needed for the three terms in the cost function. The gradient of F_{PA} can be found via the adjoint state method as described in Chapter 4 and Appendix B. For computational expediency, the gradient of F_{SUS} is computed for only a single realization of the random encoding vector $\mathbf{w}$. It is this approximation that leads to the stochastic nature of the optimization problem. Finally, for simplicity, $\mathcal{R}_c$ is chosen to be the smoothed variant of the TV semi-norm, given by Eqn. 2.36, so that the gradient can readily be computed.

Like the previously described problems for estimating the SOS distribution given by Eqns. 2.35 and 2.39, this optimization problem given by Eqn. 5.2 is non-convex. As such, the problem may have many local minima or saddle points, and a good initial guess for the SOS distribution is needed in order to avoid converging to an estimated SOS distribution that is far from the true SOS distribution. In addition to providing acoustic information to stabilize the JR approach, the USCT measurements can be employed to address this problem. An estimate of the SOS distribution can be obtained, for example, by a bent-ray image reconstruction

method, which produces lower-resolution images but which is less sensitive to the choice of the initial guess [43].

Algorithm 6 Alternating joint reconstruction method

Input: $\mathbf{c}^{(0)}, \mathbf{p}_0^{(0)}, \lambda_c, \lambda_p, \beta$
Output: $\hat{\mathbf{c}}, \hat{\mathbf{p}_0}$

$k \leftarrow 0$ {k is the algorithm iteration number.}
while stopping criterion is not satisfied **do**
  $\mathbf{p}_0^{(k+1)} \leftarrow \arg\min_{\mathbf{p}_0} F_{PA}\left(\mathbf{p}_0, \mathbf{c}^{(k)}\right) + \lambda_p \mathcal{R}_p\left(\mathbf{p}_0\right)$
  $\mathbf{c}^{(k+1)} \leftarrow \arg\min_{\mathbf{c}} F_{PA}\left(\mathbf{p}_0^{(k+1)}, \mathbf{c}\right) + \beta F_{SUS}\left(\mathbf{c}\right) + \lambda_c \mathcal{R}_c\left(\mathbf{c}\right)$
  $k \leftarrow k + 1$
end while
$\hat{\mathbf{c}} \leftarrow \mathbf{c}^{(k)}$
$\hat{\mathbf{p}_0} \leftarrow \mathbf{p}_0^{(k)}$

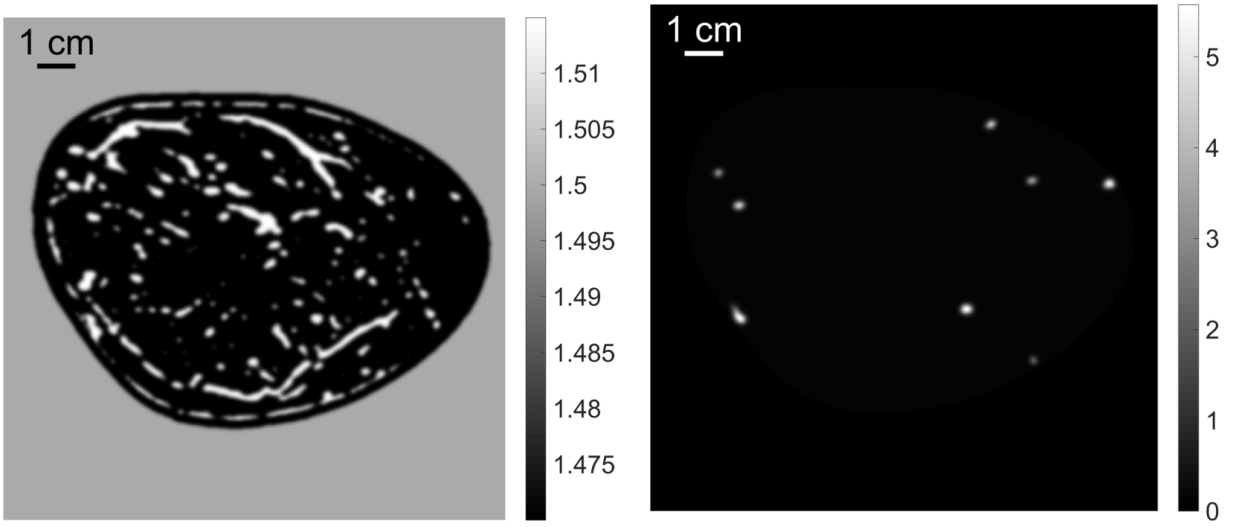
5.3 Description of computer-simulation studies

The proposed JR approach was evaluated through two-dimensional computer-simulation studies.

5.3.1 Methods

The numerical phantoms for the initial pressure and SOS distributions, shown in Fig. 5.1, were based on tissue type maps generated from segmented clinical MRI images of a human breast [67]. The SOS phantom was obtained by assigning a SOS of 1.470 mm/μs to the fatty tissue, a SOS of 1.515 mm/μs to the fibroglandular tissue, and a SOS of 1.500 mm/μs to the background medium, which is taken to be water. Similarly, the initial pressure phantom

was obtained by assigning a large initial pressure value to the blood vessels, whose cross-sections can be seen in Fig. 5.1b, a smaller non-zero initial pressure value to the parenchymal breast tissue, and an initial pressure value of zero to the background. The ratio between the maximum initial pressure value for the blood vessels and the initial pressure value for the parenchymal tissue was about 40.



(a) (b)

Figure 5.1: (a) The SOS distribution, given in units of mm/μs, and (b) the initial pressure distribution, given in arbitrary units, of the numerical breast phantom. (© 2017 IOP Publishing)

The imaging system consisted of a circular ultrasonic transducer array with 512 evenly-distributed elements and a radius of 110 mm. This geometry was chosen to be similar to existing USCT and PACT imaging systems [27,114]. The ultrasonic transducers were treated as point emitters and detectors and were placed on the simulation grid by use of nearest neighbor interpolation. The EIR of the transducers was not considered in this study.

In keeping with the studies described in Chapter 3, the excitation pulse for the USCT measurements was a windowed sine wave with a central frequency of 0.8 MHz and a length of approximately 3 cycles. When the number of emitters was less than the total number of

ultrasonic transducers, the emitters were chosen to be evenly-distributed over the circular transducer array.

In order to simulate the pressure data recorded by the transducers, the acoustic wave equations for PACT and USCT, given by Eqns. 2.21 and 2.22, respectively, were solved by use of the first-order k-space pseudo-spectral method [103]. The corresponding discrete operators, $\mathbf{H}_{PA}(\mathbf{c})$ and $\mathbf{H}_{US}(\mathbf{c})$, were implemented using NVIDIA's CUDA platform [1].

In order to avoid inverse crime [20], the measured pressure data were generated with different spatial and temporal sampling rates than were employed for image reconstruction. When the measured data were being generated, the simulation grid was a 2048×2048 uniform Cartesian grid with a pixel size of 0.125 mm, corresponding to a physical size of 256×256 mm^2. The pressure data were recorded at a temporal sampling rate of 40 MHz for 7000 time steps, corresponding to a simulation time of 175 μs. During the reconstruction, a simulation grid was a 1024×1024 uniform Cartesian grid with a pixel size of 0.25 mm, corresponding to a physical size of 256×256 mm^2. The pressure data were recorded at a temporal sampling rate of 20 MHz for 3500 time steps.

Images were reconstructed from both noiseless and noisy data. For the noisy data, additive Gaussian white noise was added to both the PACT and USCT data. For the PACT data, the noise had zero mean and a standard deviation equal to 1% of the maximum pressure recorded by the transducers. For the USCT data, the noise had zero mean and a standard deviation equal to 5% of the maximum pressure recorded by the transducer opposite the emitter for a homogeneous medium.

5.3.2 Reconstruction of the initial pressure from PACT data

Most traditional image reconstruction methods in PACT for estimating the initial pressure distribution assume a constant SOS. In order to investigate the impact of ignoring variations in the SOS, the initial pressure distribution was reconstructed from noiseless PACT data assuming two different fixed SOS distributions: the true SOS distribution and a constant SOS equal to the background SOS. In both cases, the images were reconstructed by solving Eqn. 2.31 for $\lambda = 0$. The optimization problem was solved by use of FISTA with adaptive restart and a backtracking line search [11, 12, 81]. The gradient of the data fidelity term was computed as discussed in [46]. The initial guess for the initial pressure distribution was the vector of all zeros.

5.3.3 Reconstruction of the SOS from PACT data

The SOS distribution was reconstructed from PACT data assuming two different, fixed initial pressure distributions: the true initial pressure distribution and the true initial pressure distribution shifted by 1 mm along the x-direction. The first case was performed in order to investigate the feasibility of extracting information on the SOS distribution from PACT measurements. The second was done to assess the sensitivity of the estimated SOS distribution to errors in the assumed initial pressure distribution. In both cases, the images were reconstructed from noiseless PACT data by solving Eqn. 2.35 for $\lambda = 10^{-6}$. The optimization problem was solved by use of the L-BFGS method [75, 78]. A method of computing the gradient of the data fidelity term with respect to the SOS can be found in Appendix B. A similar methodology for computing the gradient was employed in [47]. The initial guess for the SOS was a constant value equal to the background SOS.

5.3.4 Reconstruction of the SOS from USCT data

The SOS distribution can also be estimated from USCT measurements [17, 27, 39, 42, 64, 93, 113]. One key factor in determining the accuracy of the reconstructed SOS distribution in USCT is the number of views in the measured data. More views may allow the SOS distribution to be estimated more readily [24], while using less views may be decrease image acquisition times and allow for less complex and less expensive imaging hardware. Images were reconstructed from noisy USCT data for 512, 32, 16, and 8 views by solving Eqn. 2.39. For each case, the regularization parameter was swept across a range of values. The optimization problem was solved by use of SGD as described in [109]. The initial guess for the SOS was the image reconstructed by use of a bent-ray method from the corresponding USCT data [43].

5.3.5 Reconstruction of initial pressure and SOS distributions using a sequential approach

When both USCT and PACT measurements are available, the conventional approach is to reconstruct the SOS and initial pressure distributions sequentially. First, the SOS distribution is reconstructed from USCT measurements alone. Then, the initial pressure distribution is reconstructed from the PACT measurements by use of the previously estimated SOS distribution. The estimated initial pressure distributions will depend on the accuracy of the previously estimated SOS distributions. When the number of views for the USCT measurements is large, the estimated SOS distributions may be quite accurate even without utilizing the acoustic information present in the PACT measurements. As a result, we will focus on

the case where the number of USCT views is small, e.g. only 8 or 16 views. In this case, the differences between the sequential approach and the JR approach may be more pronounced.

5.3.6 Joint reconstruction from combined PACT/USCT data

Under the proposed method, the initial pressure and SOS distributions are jointly estimated from combined PACT and USCT measurements by solving Eqn. 5.1. This optimization problem is solved by use of the alternating minimization approach described by Algorithm 6. This approach involves iteratively solving two subproblems, one to estimate the initial pressure distribution and one to estimate the SOS distribution. Each of the subproblems require specification of stopping criteria. For the initial pressure subproblem, iteration was stopped when the ℓ_2-norm of the change in the object between successive iterations was less than 10^{-4} or the number of iterations exceeded 50. For the SOS subproblem, iteration was stopped when the ℓ_2-norm of the change in the object between successive iterations was less than 10^{-4} or the number of iterations exceeded 200. While these stopping criteria were found to result in accurate reconstructed SOS and initial pressure distributions, other stopping criteria could potentially accelerate the alternating JR method.

The proposed JR approach introduces a tunable parameter β (see Eqn. 5.1) that controls the relative weight of the two data fidelity terms. As the value of β is increased, the estimated SOS distribution will depend more on the information present in the USCT measurements. When β is small, the estimated SOS distribution will depend largely on the acoustic information present in the PACT measurements. In this case, the estimated SOS distribution may be similar to that obtained by JR of the SOS and initial pressure distributions from PACT measurements alone. As a result, it may be subject to the same sources of error.

When β is too large, the acoustic information present in the PACT measurements is largely ignored. To investigate the impact of β, JR of the initial pressure and SOS distributions was performed from combined noiseless PACT and USCT data for 512 USCT views for different values of β. No regularization was employed in this case as the data were noiseless and the number of USCT views was large. The initial guess for the SOS was chosen to be the image reconstructed by use of a bent-ray method from the corresponding USCT data [43].

The convergence of the alternating minimization approach was investigated for the case of noiseless data with 8 USCT views and parameter values of $\lambda_c = 1 \times 10^{-3}$, $\lambda_p = 0$, and $\beta = 10^4$. The convergence rate was evaluated in terms of the cost function values for the initial pressure and SOS subproblems of the alternating minimization approach and in terms of the root-mean-square errors (RMSEs) of the two estimated distributions. The initial guess for the SOS distribution was the image reconstructed from USCT measurements alone by use of waveform inversion for $\lambda_c = 10^{-7}$ [109].

The proposed approach was compared against the conventional sequential approach for noiseless data with 16 and 8 USCT views and noisy data with 8 USCT views. Both approaches require specification of multiple tunable parameters. While ideally the values of these parameters would be chosen to maximize performance on a specified task [10], this approach is infeasible given the high computation cost of the proposed method. Instead, these parameters were tuned to maximize the accuracy of the reconstructed images, as quantified by the RMSE. For the JR approach, the values of the parameters were chosen to minimize the RMSE of the estimated SOS distribution as it was observed that this measure also resulted in accurate recovery of the initial pressure distribution.

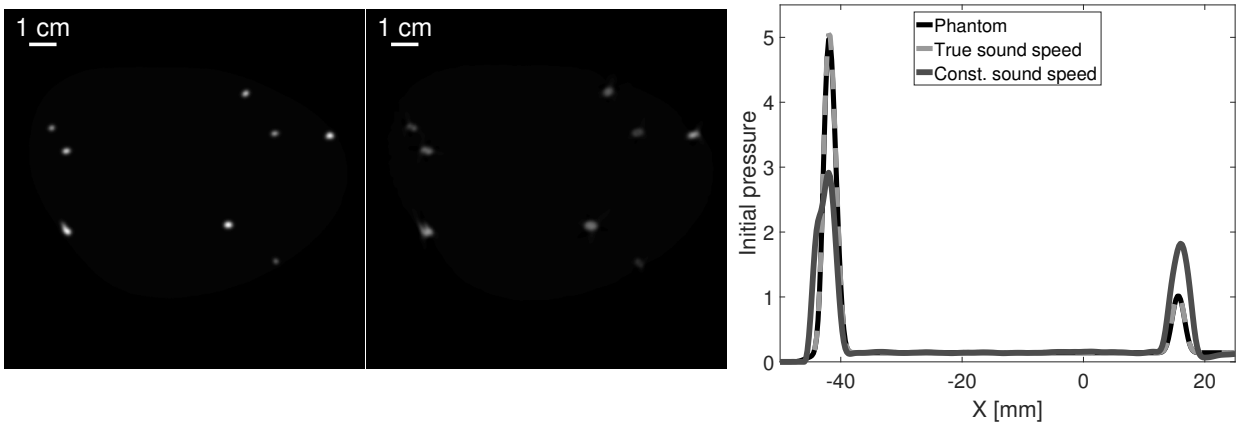
For noiseless data, tuning the regularization parameter λ_p did not lead to meaningful improvement in the accuracy of the reconstructed images. Thus, in this case, only λ_c and β

were tuned. For noisy data, the regularization parameter λ_p has a larger impact on the accuracy of the reconstructed images. To keep the number of tunable parameters manageable, the values of λ_p and λ_c were swept for $\beta = 10^3$ and for $\beta = 10^4$. These values of β were found to offer a good tradeoff between the two data fidelity terms during JR studies from noiseless data. The initial guesses for the SOS distributions were taken to be the images reconstructed from USCT measurements alone by use of waveform inversion [109].

5.4 Results of computer-simulation studies

5.4.1 Reconstruction of the initial pressure from PACT data

To study the impact of modeling variations in the SOS distribution, the initial pressure distribution was reconstructed from noiseless PACT data assuming the true SOS distribution and a constant SOS equal to the background SOS. The reconstructed images and profiles through the reconstructed images are shown in Fig. 5.2. These results confirm that ignoring the heterogeneities in the SOS distribution can result in significant errors in the reconstructed initial pressure distribution. In the initial pressure image obtained assuming a constant SOS distribution, the blood vessels are noticeably blurred and the maximum amplitude of the initial pressure distribution is underestimated. This observation is consistent with prior work demonstrating the impact of ignoring SOS variations on the reconstructed initial pressure distribution [44, 46, 51, 120].

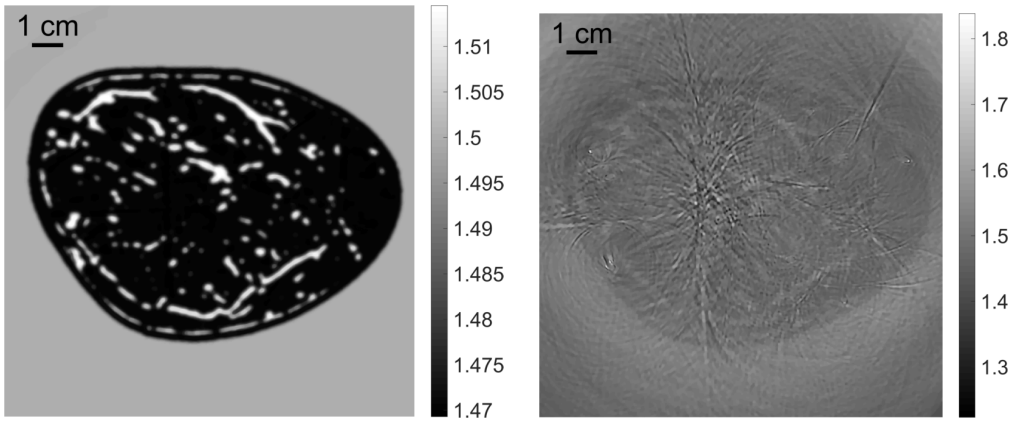


(a) (b) (c)

Figure 5.2: Initial pressure distributions reconstructed from noiseless data assuming (a) the true SOS distribution and (b) a constant SOS equal to the background SOS. (c) Profiles through the reconstructed images at $y = 14.25$ mm. Results are shown after 50 iterations. No regularization was employed. (© 2017 IOP Publishing)

5.4.2 Reconstruction of the SOS from PACT data

The SOS distribution was reconstructed from PACT data assuming two different, fixed initial pressure distributions (see Fig. 5.3). When the SOS is estimated assuming the true initial pressure distribution, the reconstructed SOS distribution is accurate. This suggests that the PACT measurements contain valuable information on the acoustic properties of the medium. However, when there are small errors in the assumed initial pressure distribution, the accuracy of the estimated SOS distribution may be much worse. This can be seen in Fig. 5.3(b), which shows the reconstructed SOS distribution when the assumed initial pressure distribution is shifted by 1 mm. This suggests the estimated SOS distribution is fairly sensitive to errors in the assumed initial pressure distribution. Thus, while the PACT measurement contain useful information on the SOS distribution, it may be difficult to extract this information via JR of the initial pressure and SOS distributions from PACT measurements alone.

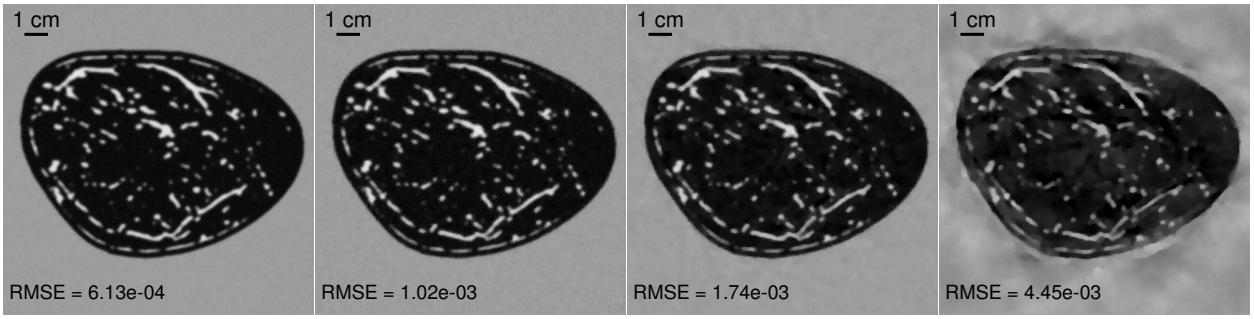


(a) (b)

Figure 5.3: Reconstructed SOS distribution from noiseless PACT data given (a) the true initial pressure distribution and (b) an initial pressure distribution shifted 1 mm in the x-direction from the true distribution. (© 2017 IOP Publishing)

5.4.3 Reconstruction of the SOS from USCT data

The SOS distribution was reconstructed from noisy USCT measurements for different numbers of views (see Fig. 5.4). The RMSE is shown in the lower left of each image. In each case, the value of the regularization parameter was tuned in order to minimize RMSE. As expected, the accuracies of the reconstructed images decrease as the number of views is decreased. The apparent visual quality of the reconstructed image is noticeably degraded for 16 views, even when the regularization parameter value is tuned. The image quality is reduced even further for the case of 8 USCT views. The obvious artifacts in the reconstructed SOS images for 8 and 16 views suggests that the accuracy of these images could be improved through the addition of PACT measurements.

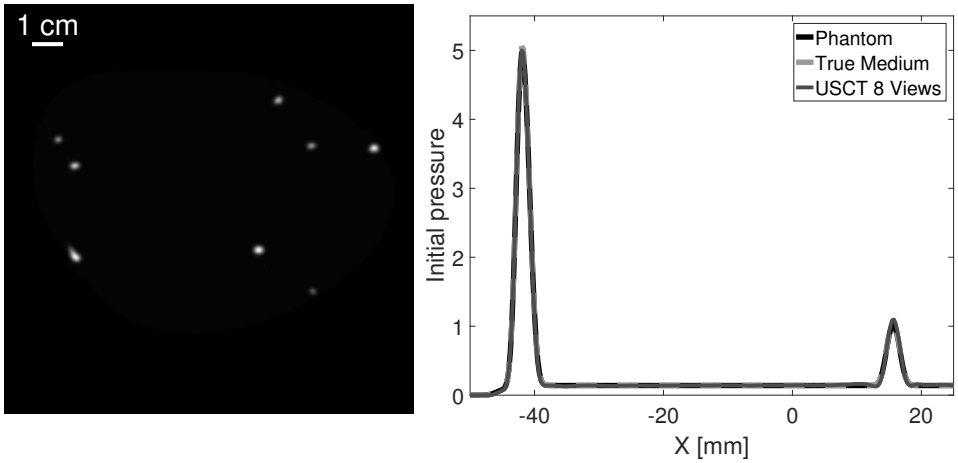


(a) (b) (c) (d)

Figure 5.4: Reconstructed SOS distributions from noisy USCT measurements for (a) 512 views and $\lambda_c = 3\times10^{-6}$, for (b) 32 views and $\lambda_c = 3\times10^{-7}$, for (c) 16 views and $\lambda_c = 3\times10^{-7}$, and for (d) 8 views and $\lambda_c = 1\times10^{-7}$. All reconstructed images are shown after 400 iterations and in a grayscale window of [1.467, 1.520] mm/μs. ()

5.4.4 Reconstruction of initial pressure and SOS distributions using a sequential approach

The conventional approach when both USCT and PACT measurements are available is to first reconstruct the SOS distribution from USCT measurements alone and then reconstruct the initial pressure distribution from the PACT measurements by use of the previously estimated SOS distribution. In this case, the estimated SOS distributions are simply those described in Section 5.4.3. As an example of the sequential approach, the initial pressure distribution was reconstructed from noiseless PACT assuming a SOS distribution equal to that obtained from noiseless USCT measurements for 8 views (see Fig. 5.5). In this case, the estimated initial pressure distribution is relatively accurately estimated even though there are small errors in the assumed SOS distribution. This suggests that the largest differences between the sequential and JR approaches may be seen in the estimated SOS distributions. In this case, the acoustic information present in the PACT measurements is the most directly impactful.



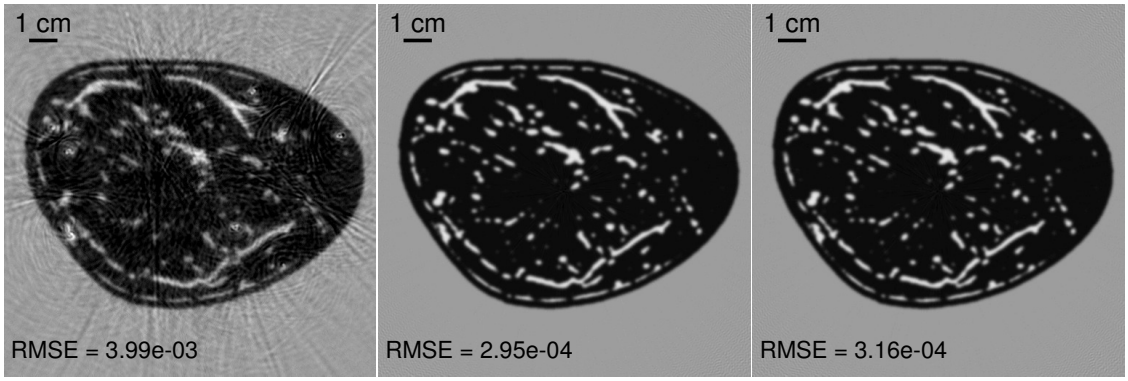
(a) (b)

Figure 5.5: (a) Reconstructed initial pressure distribution from noiseless data assuming the SOS distribution obtained from 8 noiseless USCT views. (b) Profile through the reconstructed image at $y = 14.25$ mm. The profile through the reconstructed initial pressure distribution assuming the true SOS distribution is shown as a reference. (© 2017 IOP Publishing)

5.4.5 Joint reconstruction from combined PACT/USCT data

Impact of the parameter β.

The proposed JR approach introduces a tunable parameter β (see Eqn. 5.1) that controls the relative weight of the two data fidelity terms. The reconstructed SOS distributions for different values of β can be seen in Figure 5.6. When β is too small, the estimated SOS distribution depends largely on the acoustic information present in the PACT measurements. Consistent with this, the reconstructed image for $\beta = 10^{-1}$ has artifacts that are qualitatively similar to those found in Fig. 5.3(b), which was reconstructed from PACT data alone. When β is too large, the acoustic information present in the PACT data may be largely ignored, reducing the accuracy of the estimated SOS distribution. Since the number of USCT views



(a) (b) (c)

Figure 5.6: SOS distributions reconstructed from combined noiseless PACT and USCT data for 512 USCT views for (a) $\beta = 10^{-1}$, (b) $\beta = 10^{3}$, and (c) $\beta = 10^{6}$. No regularization was employed. Results are shown after 2 outer iterations in a grayscale window of $[1.467, 1.520]$ mm/μs. (© 2017 IOP Publishing)

is large in this example, the SOS distribution can be accurately reconstructed from USCT measurements alone, and so there is only a minor increase in the RMSE when β is large.

Convergence of the alternating minimization approach.

Figure 5.7 shows the convergence of the proposed alternating minimization approach. For the chosen stopping criteria for the two subproblems, the method converges after only a few outer iterations. The estimated initial pressure distribution, in particular, changes little after only two outer iterations. This is consistent with the prior observation that the estimated initial pressure distribution is somewhat insensitive to small errors in the assumed SOS distribution [47, 82].

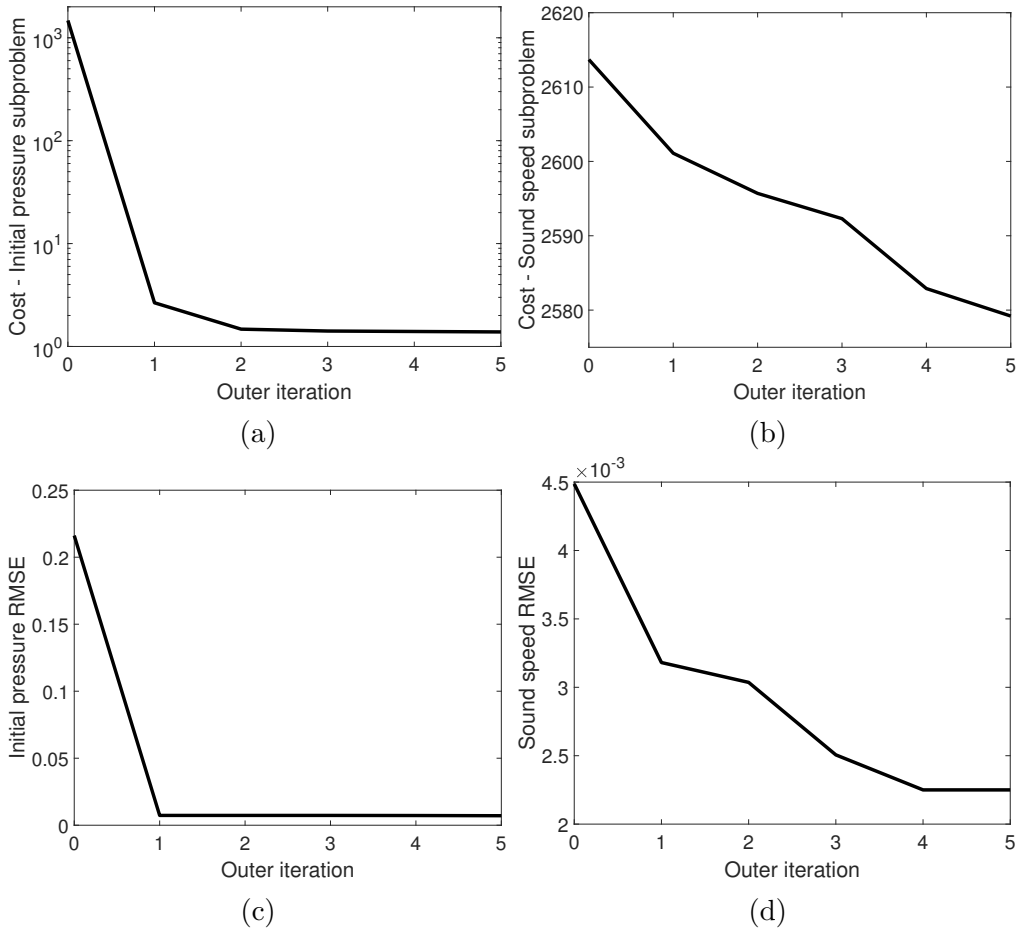


Figure 5.7: Convergence of (a) the cost function value for the initial pressure subproblem, (b) the cost function value for the SOS subproblem, (c) the initial pressure RMSE, and (d) the SOS RMSE for the case of combined noiseless PACT and USCT data for 8 USCT views. The parameters were $\lambda_c = 1 \times 10^{-3}$, $\lambda_p = 0$, and $\beta = 10^4$. For computational expediency, the value of $F_{SUS}(\mathbf{c})$ was evaluated only for a single realization of the encoding vector $\mathbf{w}$. (© 2017 IOP Publishing)

Comparison of sequential and joint reconstruction approaches.

The proposed JR approach was compared against the conventional sequential approach for noiseless data with 16 and 8 USCT views and noisy data with 8 USCT views. Recall that in the sequential approach, detailed in Section 5.3.5, the SOS distribution is estimated from USCT measurements alone. As a result, the sequential approach fails to utilize the acoustic information present in the PACT measurements, possibly leading to less accurate reconstructed SOS distributions.

The best result for the JR approach, in terms of the RMSE of the SOS, was compared with the corresponding best result obtained by use of the sequential approach for the case of noiseless data with 16 USCT views. As shown in Fig. 5.8, the JR approach is able to produce a more accurate reconstructed image for the SOS compared with the sequential approach, which only utilizes the USCT measurements to estimate the SOS. From the profiles, it can be seen that the reconstructed image for the sequential approach exhibits lower contrast and has more oscillations in the background. However, there is little difference in the reconstructed initial pressure distributions between the two approaches.

The sequential and JR approaches were also compared for the case of noiseless data corresponding to 8 USCT views. Again, the best reconstructed images, in terms of the RMSE of the SOS, for the JR approach are compared against the best reconstructed images obtained by use of the sequential approach (see Fig. 5.9). In this case, with fewer USCT views, the differences between the JR and sequential approaches are more pronounced. Still, the same general behavior is observed. The reconstructed SOS image for the sequential approach exhibits lower contrast and has more oscillations in the background. There is only a small difference in the estimated initial pressure distributions. The JR approach produced a

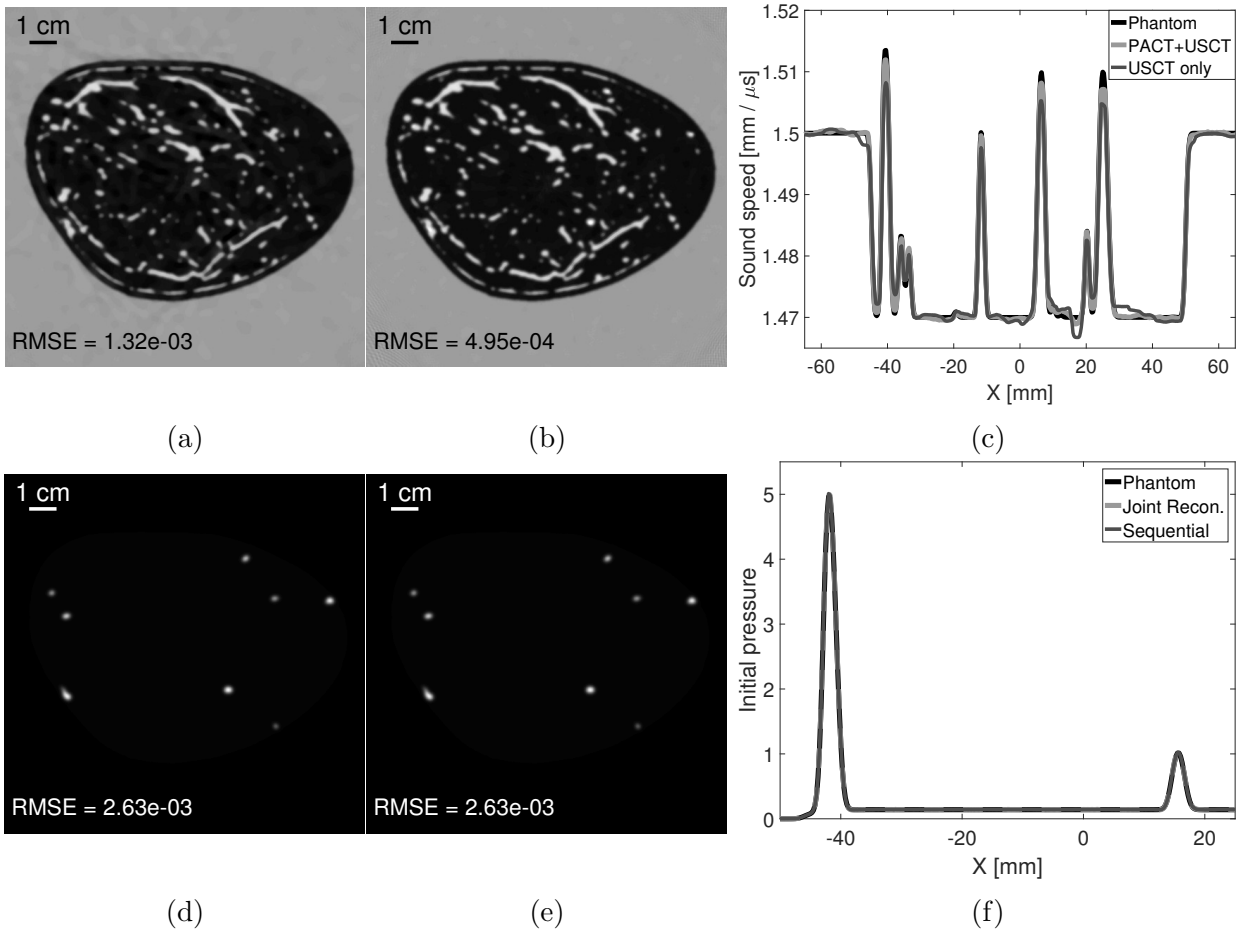


Figure 5.8: SOS distributions reconstructed from (a) noiseless USCT data and (b) noiseless PACT and USCT data for 16 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noisy data and 16 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 1 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 0$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^4$, $\lambda_c = 3 \times 10^{-4}$, and $\lambda_p = 0$. (© 2017 IOP Publishing)

slightly lower RMSE, but there is little obvious visual difference between the reconstructed initial pressure images.

Finally, the comparison was performed for the case of noisy measurements with 8 USCT views. Since the largest differences between the JR and sequential approaches were seen for the 8 USCT view case, only this case was considered here. Like before, the best reconstructed images for the JR and sequential approaches are compared. As seen in Fig. 5.10, the results are comparable to those observed in the noiseless case.

Tuning of joint reconstruction parameters

The proposed JR approach, as outlined in Algorithm 6, requires specification of several tunable parameters: β, which controls the relative weight of the USCT data fidelity term, λ_p, which controls the relative weight of the regularization term for $\mathbf{p}_0$, and λ_c, which controls the relative weight of the regularization term for $\mathbf{c}$. Tuning such a large number of parameters to maximize some measure of image quality can be a significant undertaking. Here, we briefly investigate the impact of each of these tunable parameters.

The value of the regularization parameter λ_p most directly affects the estimated initial pressure distribution. It has only an impact on the estimated SOS distribution in that the accuracy of the estimated initial pressure distribution affects how easily useful acoustic information can be extracted from the PACT data. In the case of noiseless measurements, the initial pressure distribution may be accurately estimated even when $\lambda_p = 0$. Thus, for noiseless data, only β and λ_c are tuned. However, these two parameters must be tuned jointly. Consider that as β is increased, not only is the relative weight of the USCT data fidelity term increased relative to the PACT data fidelity term, but also the relative weight

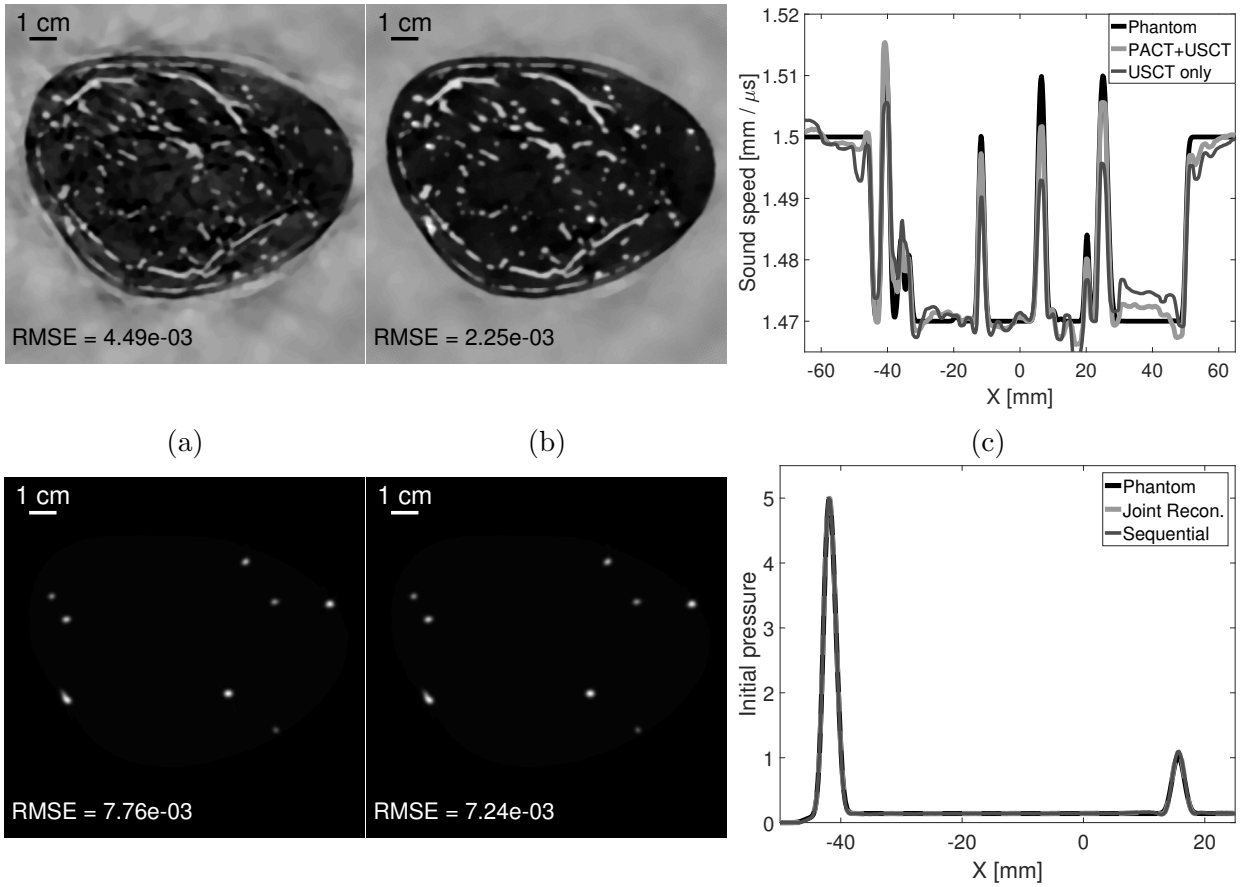


Figure 5.9: SOS distributions reconstructed from (a) noiseless USCT data and (b) noiseless PACT and USCT data for 8 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noiseless data and 8 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 1 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 0$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^4$, $\lambda_c = 1 \times 10^{-3}$, and $\lambda_p = 0$. (© 2017 IOP Publishing)

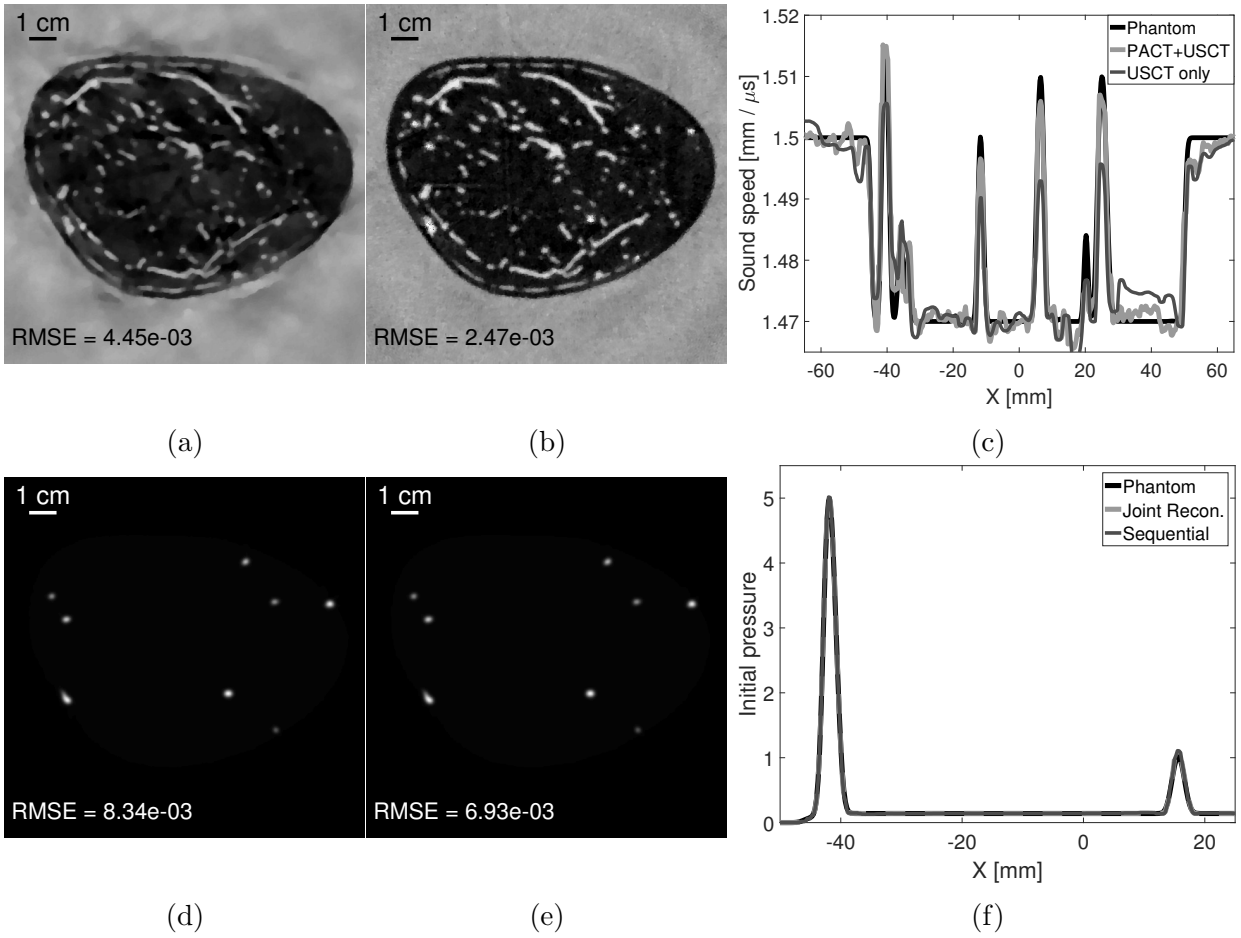


Figure 5.10: SOS distributions reconstructed from (a) noisy USCT data and (b) noisy PACT and USCT data for 8 USCT views. (c) Profiles through the reconstructed images at y = 16.75 mm. Initial pressure distributions reconstructed by (d) the sequential approach and (e) the JR approach for noisy data and 8 USCT views. (f) Profiles through the reconstructed images at y = 14.25 mm. The sequential SOS distribution is shown for $\lambda_c = 3 \times 10^{-7}$ after 400 iterations. The sequential initial pressure distribution is shown for $\lambda_p = 3 \times 10^{-3}$ after 50 iterations. The JR results correspond to 4 outer iterations for $\beta = 10^3$, $\lambda_c = 1 \times 10^{-4}$, and $\lambda_p = 1 \times 10^{-2}$. (© 2017 IOP Publishing)

of the USCT data fidelity term is increased relative to the SOS regularization term. Thus, the optimal value of λ_c according to some measure may be different for different values of β. For noisy measurements, all three parameters must be jointly tuned.

As summary measures, the RMSEs of the reconstructed SOS images obtained by tuning β and λ_c for noiseless measurements for 16 USCT views are given in Table 5.1. The RMSEs of the reconstructed images obtained by JR are reduced relative to the RMSEs obtained for images reconstructed from USCT measurements alone for a wide range of values of β and λ_c. This suggests that the proposed JR approach is not overly sensitive to the three parameters and some improvement in the accuracy of the reconstructed images can be achieved without precise tuning the values of β and λ_c. As a further example, the RMSEs of the reconstructed SOS images for tuning β and λ_c for noiseless measurements for 8 USCT views are summarized in Table 5.2. The results similarly show an improvement in RMSEs obtained by JR compared with those obtained from USCT measurements alone for a wide range of values of β and λ_c.

For noisy measurements, all three parameters must be tuned. For computationally expediency, results are shown for only a couple of values of β. For the case of noisy measurements for 8 USCT views, λ_p and λ_c are jointly tuned for $\beta = 10^3$ and $\beta = 10^4$. The RMSEs for the reconstructed SOS images for the two approaches are summarized in Table 5.3. In this case, changing the value of λ_p does impact the accuracy of the reconstructed SOS. However, the impact is predictably less significant than that of λ_c. As in noiseless case, there is an improvement in the RMSE of the reconstructed images compared with those for USCT measurements alone across a wide range of values of the parameters.

Table 5.1: RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noiseless data for 16 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing)

	λ_c								
β	3E-6	1E-5	3E-5	1E-4	3E-4	1E-3	3E-3	1E-2	1E-2
10^2	1.18	1.14	1.15	1.52	–	–	–	–	–
10^3	–	0.87	0.84	0.86	1.14	–	–	–	–
10^4	–	–	0.59	0.57	**0.50**	0.64	–	–	–
10^5	–	–	–	–	–	0.56	0.52	0.60	1.09

Table 5.2: RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noiseless data for 8 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing)

	λ_c							
β	1E-5	3E-5	1E-4	3E-4	1E-3	3E-3	1E-2	3E-2
10^2	3.11	2.99	3.05	4.06	–	–	–	–
10^3	2.62	2.52	2.40	2.65	–	–	–	–
10^4	–	–	2.45	2.27	**2.25**	2.71	–	–
10^5	–	–	–	–	3.22	3.02	2.65	2.94

Table 5.3: RMSEs ($\times 10^{-3}$) of the reconstructed SOS images for JR from noisy data for 8 USCT views for different parameter values. The lowest RMSE value is shown in bold. (© 2017 IOP Publishing)

$\beta = 10^3$				$\beta = 10^4$			
	λ_c				λ_c		
λ_p	1E-5	1E-4	1E-3	λ_p	1E-4	1E-3	1E-2
3E-3	6.21	2.61	3.70	1E-2	5.56	2.92	4.32
1E-2	5.42	**2.47**	3.78	3E-2	5.53	2.87	4.31
3E-2	6.84	2.79	3.54	1E-1	5.63	2.92	4.33

5.5 Conclusions

For the first time, JR of the initial pressure and SOS distributions was performed from combined PACT/USCT measurements, enabling the acoustic information in the PACT data to be effectively utilized. Exploiting this information allows both the initial pressure and the SOS distributions to be more accurately estimated. Compared with use of the sequential approach, which estimates the SOS distribution from USCT measurements alone, the JR approach allows the SOS distribution to be more accurately estimated from few USCT views. In turn, the initial pressure distribution may be more accurately reconstructed due to the improved estimate of the SOS distribution. By requiring only few-view USCT data, the proposed approach could allow for construction of simpler and lower cost combined PACT/USCT imaging systems that require fewer transducers to act as emitters. Additionally, image acquisition times could be reduced. By jointly considering combined PACT/USCT measurement data, a synergistic improvement in the reconstructed image quality is observed. The initial pressure distribution is more accurately estimated by the use of USCT measurements and the SOS distribution is more accurately estimated by the use of PACT measurements.

The proposed approach, however, does have some limitations. First, the method requires specification of a number of tunable parameters, in particular β, λ_p, and λ_c. Choosing these parameters to maximize image quality can require time-consuming manual tuning. Second, the method is computationally expensive. Each outer iteration of the JR approach requires solving two subproblems, one to estimate the SOS distribution and one to estimate the initial pressure distribution. The subproblem for estimating the SOS, in particular, is very computationally demanding. Fortunately, only a few outer iterations are needed to reach convergence. Third, the imaging models employed for these initial studies ignore several

aspects of experimental imaging systems. In particular, the assumed models simulate 2-D acoustic wave propagation rather than the 3-D propagation that occurs experimentally. Further, density variations, acoustic attenuation, and the spatial and electro-acoustic impulse responses of the ultrasonic transducers are ignored.

Chapter 6

Summary

In this dissertation, we have developed methods for image reconstruction in USCT and PACT. These approaches sought to obtain accurate high-resolution images by employing imaging models based on the acoustic wave equation and incorporating these models within an optimization-based image reconstruction framework.

In USCT, a new stochastic optimization method was proposed for reconstruction of the SOS distribution. This approach built on previously developed source encoding methods to greatly reduce the computational cost of waveform inversion. By use of the RDA method, three key shortcomings of previous methods were overcome. First, the proposed method allows use of non-smooth regularization functions that permit the known sparseness properties of the object to be exploited during image reconstruction. Second, the RDA method employs knowledge of the structure of the cost function, namely the stochastic nature of the data fidelity term and the deterministic nature of the regularization term, to achieve more effective regularization that offers a superior tradeoff between noise and bias. Third, the approach allows use of a line search to improve convergence rates without reducing accuracy due to overfitting.

In PACT, two approaches were proposed to compensate for the unknown SOS distribution. In the first, the initial pressure and SOS distributions were jointly estimated from PACT data alone by use of a low-dimensional parameterized model for the SOS. By use of a low-dimensional model, the instability of the general (unparameterized) JR problem can be mitigated. The detailed approach allowed the gradients with respect to both the initial pressure and SOS distributions to be calculated with only two wave solver runs. It also permitted use of non-smooth regularization functions and constraints for the initial pressure distribution and was capable of accurately modeling objects with large variations in the SOS distribution (for example, like those found between soft tissue and bone). In the second approach, the initial pressure and SOS distributions were jointly estimated from combined PACT/USCT measurements. By exploiting the acoustic information present in the PACT measurements, few USCT views were needed to accurately reconstruct the SOS distribution. Additionally, the reconstructed initial pressure distributions were more accurate than those obtained by assuming a constant SOS.

Many areas remain for further improvement. While the chosen imaging models sought to more accurately model the underlying physics of the imaging system, they still have several shortcomings. For example, the models all consider 2D acoustic propagation while the propagation for experimental systems is inherently 3D. They also ignore mass density variations, acoustic absorption and dispersion, and do not model the spatial impulse responses of the ultrasonic transducers. The JR problems could be extended to other imaging contrasts such as the density or shear properties of the medium. This could be useful for transcranial PACT where the acoustic properties of the skull are typically unknown and have a transformative impact on the recorded pressure signals [74,89]. Further, acceleration of the proposed approaches, in particular the JR approaches, could help increase their attractiveness.

Appendix A

Adjoint state method

Many optimization methods require an efficient method to calculate the gradient of the objective function with respect to the sought-after parameters in order to be practical. While finite-difference-based methods provide a straight-forward approach to calculating the gradient, this approach requires the objective function to be evaluated a number of times that scales linearly with the number of parameters [36]. This can make the approach infeasible for large-scale optimization problems.

For linear least squares objective functions, such as

$$F(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2, \tag{A.1}$$

the gradient can be calculated simply by use of matrix calculus [34]. When $\mathbf{A}$ is a real-valued matrix, the gradient is given by

$$\nabla_{\mathbf{x}} F = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{y}), \tag{A.2}$$

where the subscript ^{T} denotes the transpose. In this case, assuming the computational cost is dominated by the application of $\mathbf{A}$, the cost of calculating the gradient is roughly equivalent to evaluating the objective function twice (and therefore is independent of the number of parameters). When estimating the initial pressure distribution in PACT, the data fidelity term falls into this category. The application of this approach for forward models based on the acoustic wave equation and the elastic wave equation have previously been described by Huang, et al. [46] and Mitsuhashi, et al. [74], respectively.

When the model is nonlinear with respect to the sought-after parameters, the above approach cannot be employed. Thankfully, there is an alternative method, known as the adjoint state method, that allows the gradient of an arbitrary functional to be efficiently computed. This approach has seen widespread application in number of different fields. The backpropagation algorithm employed in machine learning arises from this approach, for example [63]. The presentation, here, follows the excellent reviews found in the inverse scattering and geophysics literature [79, 87]. First, a general overview of the adjoint state method is provided. Then, in Appendix B, an example of applying the method for joint reconstruction in PACT is given. While the adjoint state method can be derived using several different approaches, here, the Lagrangian interpretation is given. For simplicity, the presentation is performed in a continuous setting.

The goal of the adjoint state method is to calculate the gradient of a functional F with respect to some model parameters m. Let us assume the functional depends on the model parameters, in part, through some state variables Ψ,

$$F(m) = f(\Psi(m), m). \tag{A.3}$$

For example, m could be the SOS or initial pressure distributions, Ψ could be the acoustic pressure, and F could be the least squares error between the measured acoustic pressure and the estimated acoustic pressure given some model. This model defines the relationship between m and Ψ. As will be seen, it can be useful to denote this model by a collection of K constraints,

$$h_k\left(\Psi, m\right) = 0, \tag{A.4}$$

for $k \in [0,\ K-1]$. These constraints could correspond to the acoustic wave equation, given by Eqn. 2.18.

The derivative of F with respect to m can be calculated via the chain rule as

$$\frac{dF}{dm} = \frac{\partial f}{\partial \Psi}\frac{\partial \Psi}{\partial m} + \frac{\partial f}{\partial m}. \tag{A.5}$$

Direct application of this expression, however, seems to indicate that calculation of $\frac{dF}{dm}$ requires calculation of $\frac{\partial \Psi}{\partial m}$, which is often high-dimensional and expensive to compute.

A key benefit of the adjoint state method is that it provides a prescription for calculating $\frac{dF}{dm}$ without explicitly computing $\frac{\partial \Psi}{\partial m}$. Consider the augmented functional $\mathcal{L}$, which incorporates the model-based constraints, given by

$$\mathcal{L}\left(\Psi, m, \lambda\right) = f\left(\Psi, m\right) - \sum_k \left\langle \lambda_k, h_k\left(\Psi, m\right)\right\rangle, \tag{A.6}$$

where λ_k are the adjoint state variables and λ is shorthand for the collection of all K adjoint state variables. This expression is actually the Lagrangian for the optimization problem of minimizing f with respect to Ψ subject to the constraints $h_k\left(\Psi, m\right) = 0$. When Ψ and m

are related by these constraints,

$$\frac{dF}{dm} = \frac{\partial \mathcal{L}}{\partial \Psi}\frac{\partial \Psi}{\partial m} + \frac{\partial \mathcal{L}}{\partial m}. \tag{A.7}$$

This holds independent of the choice of λ. In order to allow Eqn. A.7 to be evaluated efficiently, λ is chosen such that

$$\frac{\partial \mathcal{L}}{\partial \Psi} = \frac{\partial f}{\partial \Psi} - \sum_k \frac{\partial}{\partial \Psi}\left\langle \lambda_k, h_k \right\rangle = 0. \tag{A.8}$$

This is known as the adjoint state equation. Let $\widetilde{\lambda}$ denote the solution to Eqn. A.8 and let $\widetilde{\Psi}$ satisfy the constraints $h_k\left(\Psi, m\right) = 0$. Then, $\left(\widetilde{\Psi}, \widetilde{\lambda}\right)$ represents a saddle point of $\mathcal{L}$ since $\frac{\partial \mathcal{L}}{\partial \Psi} = 0$ and $\frac{\partial \mathcal{L}}{\partial \lambda_k} = -h_k\left(\Psi, m\right) = 0$ at that point. Once $\widetilde{\Psi}$ and $\widetilde{\lambda}$ are known, the derivative of F with respect to m can finally be calculated as,

$$\frac{dF}{dm} = \frac{\partial \mathcal{L}\left(\widetilde{\Psi}, m, \widetilde{\lambda}\right)}{\partial m} = \frac{\partial f}{\partial m} - \sum_k \left\langle \widetilde{\lambda_k}, \frac{\partial h_k}{\partial m} \right\rangle. \tag{A.9}$$

Thus, the adjoint state method follows a three-part prescription. First, calculate the state variable $\widetilde{\Psi}$ consistent with m and the chosen model. Second, compute of the value of the adjoint state variables $\widetilde{\lambda}$. Third, compute the derivative of the cost function by use of $\widetilde{\Psi}$ and $\widetilde{\lambda}$. While the adjoint state method has broad applicability, several conditions must be met for the method to be employed in practice: (1) F, f, h_k and Ψ must be continuously differentiable; (2) each m must correspond to a unique Ψ; and (3) there must exist an efficient method for solving Eqn. A.8.

Appendix B

Applications of the adjoint state method

In Appendix A, a general, and somewhat abstract, overview of the adjoint state method was provided. In this section, the adjoint state method is employed to compute the gradients of the data fidelity term given in Eqn. 2.32 with respect to the initial pressure and SOS distributions. The adjoint state method has previously been employed to calculate the gradients of data fidelity terms in PACT and USCT with respect to the SOS distribution [47, 109].

For simplicity, the presentation focuses on the first-order acoustic wave for PACT given in Eqn. 2.21. The approaches for USCT and for the second-order acoustic wave equation follow the same basic method and have been detailed previously [79, 87, 109]. When comparing expressions for first-order and second-order acoustic wave equations, keep in mind that the definitions of the time-varying source terms differ (see Eqns. 2.22 and 2.23). Additionally, integration-by-parts allows the temporal derivatives to be moved about the expression for the gradient yielding several seemingly distinct yet equivalent expressions.

To start, consider the constraints correspond to the acoustic wave equation given by Eqn. 2.18,

$$h_0\left(p, \mathbf{u}, p_0, c\right) = \rho\left(\mathbf{r}\right) \frac{\partial \mathbf{u}\left(\mathbf{r}, t\right)}{\partial t} + \nabla p\left(\mathbf{r}, t\right) \tag{B.1a}$$

$$h_1\left(p, \mathbf{u}, p_0, c\right) = \frac{1}{\rho\left(\mathbf{r}\right) c\left(\mathbf{r}\right)^2} \frac{\partial p\left(\mathbf{r}, t\right)}{\partial t} + \nabla \cdot \mathbf{u}\left(\mathbf{r}, t\right) \tag{B.1b}$$

$$h_2\left(p, \mathbf{u}, p_0, c\right) = p\left(\mathbf{r}, 0\right) - p_0\left(\mathbf{r}\right) \tag{B.1c}$$

$$h_3\left(p, \mathbf{u}, p_0, c\right) = \mathbf{u}\left(\mathbf{r}, 0\right), \tag{B.1d}$$

where p and $\mathbf{u}$ are the state variables and c and p_0 are the sought-after model parameters. A method for computing the values of the acoustic pressure and particle velocity given the SOS and initial pressure distributions, i.e. a method for solving the acoustic wave equation, was previously discussed in Section 2.1.4. Thus, we shall move on to the task of determining the adjoint state variables for this model.

To do this, it is necessary to define the functional to be minimized, which is taken to be

$$f\left(p\right) = \frac{1}{2} \sum_{i=0}^{M-1} \int_0^T dt \left(\underline{g_i}\left(t\right) - \mathcal{M}_i p\left(\mathbf{r}, t\right)\right)^2 \tag{B.2}$$

where $\underline{g_i}\left(t\right)$ is the pressure recorded by the i-th transducer and $\mathcal{M}_i$ is the restriction of the pressure over the whole domain to the location of the i-th transducer. Note that this function depends only on the acoustic pressure and not directly on the particle velocity. The

corresponding augmented functional is given by

$$\begin{aligned}\mathcal{L} = & \frac{1}{2}\sum_{i=0}^{M-1}\int_0^T dt\left(g_i\left(t\right) - \mathcal{M}_i p\left(\mathbf{r},t\right)\right)^2 \\ & - \int_0^T dt\left\langle \boldsymbol{\lambda}_1, \rho\frac{\partial \mathbf{u}}{\partial t} + \nabla p\right\rangle_U \\ & - \int_0^T dt\left\langle \lambda_2, \frac{1}{\rho c^2}\frac{\partial p}{\partial t} + \nabla\cdot\mathbf{u}\right\rangle_V \\ & - \left\langle \boldsymbol{\lambda}_3, \mathbf{u}\left(\mathbf{r},0\right)\right\rangle_U - \left\langle \lambda_4, p\left(\mathbf{r},0\right) - p_0\left(\mathbf{r}\right)\right\rangle_V,\end{aligned} \tag{B.3}$$

where

$$\left\langle \mathbf{x},\mathbf{y}\right\rangle_U = \sum_{i=1}^{n}\int d\mathbf{r}\; x_i\left(\mathbf{r}\right) y_i\left(\mathbf{r}\right) \quad \text{and} \quad \left\langle a,b\right\rangle_V = \int d\mathbf{r}\; a\left(\mathbf{r}\right) b\left(\mathbf{r}\right)$$

are the inner products for vector-valued and scalar-valued quantities, respectively. Here, n is the number of components in the vector-valued quantity and x_i represents the i-th component of $\mathbf{x}$.

To compute the derivatives needed to solve the adjoint state equations defined by Eqn. A.8, it is useful to rearrange the constraint terms. This is done by use of integration-by-parts and by the fact that the adjoint of the gradient operator is the negative divergence operator. With this, the term in the Lagrangian corresponding to Eqn. B.1a can be rewritten as

$$\begin{aligned}&\int_0^T dt\left\langle \boldsymbol{\lambda}_1, \rho\frac{\partial \mathbf{u}}{\partial t} + \nabla p\right\rangle_U \\ &\quad = \int_0^T dt\left[\left\langle \boldsymbol{\lambda}_1, \rho\frac{\partial \mathbf{u}}{\partial t}\right\rangle_U + \left\langle \boldsymbol{\lambda}_1, \nabla p\right\rangle_U\right] \\ &\quad = \int_0^T dt\left[\left\langle \rho\boldsymbol{\lambda}_1, \frac{\partial \mathbf{u}}{\partial t}\right\rangle_U - \left\langle \nabla\cdot\boldsymbol{\lambda}_1, p\right\rangle_V\right] \\ &\quad = \left\langle \rho\boldsymbol{\lambda}_1, \mathbf{u}\right\rangle_U\Big|_0^T - \int_0^T dt\left[\left\langle \rho\frac{\partial \boldsymbol{\lambda}_1}{\partial t}, \mathbf{u}\right\rangle_U + \left\langle \nabla\cdot\boldsymbol{\lambda}_1, p\right\rangle_V\right].\end{aligned}$$

Similarly, the term in the Lagrangian corresponding to Eqn. B.1b can be rewritten as

$$\begin{aligned}
&\int_0^T dt \left\langle \lambda_2, \frac{1}{\rho c^2}\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{u} \right\rangle_V \\
&\quad = \int_0^T dt \left[\left\langle \lambda_2, \frac{1}{\rho c^2}\frac{\partial p}{\partial t} \right\rangle + \langle \lambda_2, \nabla \cdot \mathbf{u} \rangle_V \right] \\
&\quad = \int_0^T dt \left[\left\langle \frac{1}{\rho c^2}\lambda_2, \frac{\partial p}{\partial t} \right\rangle_V - \langle \nabla \lambda_2, \mathbf{u} \rangle_U \right] \\
&\quad = \left\langle \frac{1}{\rho c^2}\lambda_2, p \right\rangle_V \Bigg|_0^T - \\
&\qquad \int_0^T dt \left[\left\langle \frac{1}{\rho c^2}\frac{\partial \lambda_2}{\partial t}, p \right\rangle_V + \langle \nabla \lambda_2, \mathbf{u} \rangle_U \right].
\end{aligned}$$

Finally, the overall Lagrangian can be rewritten as

$$\begin{aligned}
\mathcal{L} = &\frac{1}{2}\sum_{i=0}^{M-1} \int_0^T dt \left(g_i(t) - \mathcal{M}_i p(\mathbf{r}, t) \right)^2 \\
&- \langle \rho \boldsymbol{\lambda}_1, \mathbf{u} \rangle_U \,|_0^T + \int_0^T dt \left[\left\langle \rho \frac{\partial \boldsymbol{\lambda}_1}{\partial t}, \mathbf{u} \right\rangle_U + \langle \nabla \cdot \boldsymbol{\lambda}_1, p \rangle_V \right] \\
&- \left\langle \frac{1}{\rho c^2}\lambda_2, p \right\rangle_V \Bigg|_0^T + \int_0^T dt \left[\left\langle \frac{1}{\rho c^2}\frac{\partial \lambda_2}{\partial t}, p \right\rangle_V + \langle \nabla \lambda_2, \mathbf{u} \rangle_U \right] \\
&- \langle \boldsymbol{\lambda}_3, \mathbf{u}(\mathbf{r}, 0) \rangle_U - \langle \lambda_4, p(\mathbf{r}, 0) - p_0(\mathbf{r}) \rangle_V .
\end{aligned} \tag{B.4}$$

Now the derivatives of the Lagrangian with respect to the state variables can be readily computed. To obtain an expression for the adjoint state equation, compute $\frac{\partial \mathcal{L}}{\partial p}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{u}}$ and set the expressions equal to zero. Then, group the terms by their time dependence. This

yields

$$\rho(\mathbf{r})\frac{\partial \boldsymbol{\lambda}_1(\mathbf{r},t)}{\partial t}+\nabla\lambda_2(\mathbf{r},t)=\mathbf{0} \tag{B.5a}$$

$$\nabla\cdot\boldsymbol{\lambda}_1(\mathbf{r},t)+\frac{1}{\rho(\mathbf{r})\,c(\mathbf{r})^2}\frac{\partial\lambda_2(\mathbf{r},t)}{\partial t}=-\tilde{S}(\mathbf{r},t) \tag{B.5b}$$

$$\boldsymbol{\lambda}_1(\mathbf{r},T)=0 \tag{B.5c}$$

$$\lambda_2(\mathbf{r},T)=0 \tag{B.5d}$$

$$\boldsymbol{\lambda}_3(\mathbf{r})=\rho(\mathbf{r})\,\boldsymbol{\lambda}_1(\mathbf{r},0) \tag{B.5e}$$

$$\lambda_4(\mathbf{r})=\frac{1}{\rho(\mathbf{r})\,c(\mathbf{r})^2}\lambda_2(\mathbf{r},0)\,, \tag{B.5f}$$

where

$$\tilde{S}(\mathbf{r},t)=\sum_{i=0}^{M-1}\mathcal{M}_i^{\dagger}\left(\mathcal{M}_i p(\mathbf{r},t)-\underline{g_i}(t)\right). \tag{B.6}$$

The superscript † is employed to denote the adjoint. The action of $\mathcal{M}_i^{\dagger}$ can be computed as

$$\mathcal{M}_i^{\dagger}g(t)=\Lambda_i^{\dagger}\left(h_e^{\dagger}(t)*_t g(t)\right), \tag{B.7}$$

where the action of $\Lambda_i^{\dagger}$ can be computed by placing the corresponding value at the location of the i-th transducer. Further, the temporal convolution can be computed as

$$\begin{aligned}h_e^{\dagger}(t)*_t g(t)&=h_e(-t)*_t g(t)\\&=\mathcal{F}^{-1}\left\{\mathcal{F}\left\{h_e\right\}^{*}\mathcal{F}\left\{g\right\}\right\},\end{aligned}$$

where $\mathcal{F}$ is the Fourier transform and * denotes the complex conjugate.

The set of differential equations given by Eqn. B.5 bear a close resemblance to the first-order acoustic wave equation given by Eqn. 2.18. This suggests that a solution to these differential equations could possibly be computed by a similar approach. However, one difficulty is that Eqn. B.5 involves final conditions on $\boldsymbol{\lambda}_1$ and λ_2 rather than the initial conditions specified in Eqn. 2.18. A change of variables can be employed to change the final conditions to initial conditions:

$$q_i(\mathbf{r}, t) = \lambda_i(\mathbf{r}, T - t). \tag{B.8}$$

This gives a set of differential equations for the adjoint state variables that can be solved by the same method as employed to solve Eqn. 2.18:

$$\rho(\mathbf{r}) \frac{\partial \mathbf{q}_1(\mathbf{r}, t)}{\partial t} + \nabla q_2(\mathbf{r}, t) = \mathbf{0} \tag{B.9a}$$

$$\nabla \cdot \mathbf{q}_1(\mathbf{r}, t) + \frac{1}{\rho(\mathbf{r}) c(\mathbf{r})^2} \frac{\partial q_2(\mathbf{r}, t)}{\partial t} = -\tilde{S}(\mathbf{r}, T - t) \tag{B.9b}$$

$$\mathbf{q}_1(\mathbf{r}, 0) = 0 \tag{B.9c}$$

$$q_2(\mathbf{r}, 0) = 0. \tag{B.9d}$$

As such, the computational cost of calculating the adjoint state variables is the same is solving the acoustic wave equation for the forward problem.

The last step is to compute the gradients with respect to the model parameters by use of the previously calculated state and adjoint state variables. This can be done as

$$\frac{dF}{dp_0}(\mathbf{r}) = \lambda_4(\mathbf{r}) = \frac{1}{\rho(\mathbf{r}) c(\mathbf{r})^2} q_2(\mathbf{r}, T) \tag{B.10}$$

$$\frac{dF}{dc}(\mathbf{r}) = -\int_0^T dt \frac{2}{\rho(\mathbf{r}) c(\mathbf{r})^3} \frac{\partial p(\mathbf{r}, t)}{\partial t} q_2(\mathbf{r}, T - t). \tag{B.11}$$

In a discrete setting, the above quantities may be replaced by their corresponding discrete approximations (see Section 2.1.3).

References

[1] CUDA C Programming Guide. Technical Report PG-02829-001_v7.5, NVIDIA Corporation, September 2015.

[2] H. Abdullah and A.K. Louis. The approximate inverse for solving an inverse scattering problem for acoustic waves in an inhomogeneous medium. *Inverse Problems*, 15:1213–1229, 1999.

[3] M.A. Anastasio, Jin Zhang, E.Y. Sidky, Yu Zou, D. Xia, and Xiaochuan Pan. Feasibility of half-data image reconstruction in 3-D reflectivity tomography with a spherical aperture. *IEEE Transactions on Medical Imaging*, 24(9):1100–1112, September 2005.

[4] M.A. Anastasio, Jin Zhang, Xiaochuan Pan, Yu Zou, Geng Ku, and L.V. Wang. Half-time image reconstruction in thermoacoustic tomography. *IEEE Transactions on Medical Imaging*, 24(2):199–210, February 2005.

[5] Michael P. Andr, Helmar S. Jane, Peter J. Martin, Gregory P. Otto, Brett A. Spivey, and Douglas A. Palmer. High-speed data acquisition in a diffraction tomography system employing large-scale toroidal arrays. *International Journal of Imaging Systems and Technology*, 8(1):137–147, 1997.

[6] Fatima Anis, Yang Lou, Andr Conjusteau, Richard Su, Tanmayi Oruganti, Sergey A. Ermilov, Alexander A. Oraevsky, and Mark A. Anastasio. Investigation of the adjoint-state method for ultrasound computed tomography: a numerical and experimental study. In *Proc. SPIE*, volume 8943, page 894337, March 2014.

[7] Simon R. Arridge, Marta M. Betcke, Ben T. Cox, Felix Lucka, and Brad E. Treeby. On the Adjoint Operator in Photoacoustic Tomography. *Inverse Problems*, 32:115012, 2016.

[8] Haim Avron and Sivan Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of ACM*, 58(2):8:1–8:17, 2011.

[9] R. G. Baraniuk. Compressive Sensing [Lecture Notes]. *IEEE Signal Processing Magazine*, 24(4):118–121, July 2007.

[10] Harrison H. Barrett and Kyle J. Myers. *Foundations of Image Science*. Wiley series in pure and applied optics. Wiley-Interscience, Hoboken, NJ, 2004.

[11] Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[12] Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, January 2009.

[13] Jean-Pierre Berenger. A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, 1994.

[14] Simon Bernard, Vadim Monteiller, Dimitri Komatitsch, and Philippe Lasaygues. Ultrasonic computed tomography based on full-waveform inversion for bone quantitative imaging. *Physics in Medicine & Biology*, 62(17):7011, 2017.

[15] Junguo Bian, Jeffrey H Siewerdsen, Xiao Han, Emil Y Sidky, Jerry L Prince, Charles A Pelizzari, and Xiaochuan Pan. Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Physics in Medicine and Biology*, 55(22):6575–6599, November 2010.

[16] C. Bunks, F. Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, September 1995.

[17] P. L. Carson, C. R. Meyer, A. L. Scherzinger, and T. V. Oughton. Breast imaging in coronal planes with simultaneous pulse echo and transmission ultrasound. *Science*, 214(4525):1141–1143, December 1981.

[18] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock. An introduction to total variation for image analysis. In Massino Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, number 9 in Radon Series on Computational and Applied Mathematics, pages 263–340. de Gruyter, 2010.

[19] Antonin Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004.

[20] David L. Colton and Rainer Kress. *Inverse acoustic and electromagnetic scattering theory*. Springer-Verlag, New York, 3rd edition, 2013.

[21] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In Heinz H. Bauschke, R. Burachik, P. Combettes, V. Elser, D. Luke, and H. Wolkowicz, editors, *Fixed-point algorithms for inverse problems in science and engineering*. Springer, New York, NY, 2011.

[22] Ben Cox. Acoustics for Ultrasound Imaging. *Lecture Notes, University College London*, 2012.

[23] Ben Cox, Jan G. Laufer, Simon R. Arridge, and Paul C. Beard. Quantitative spectroscopic photoacoustic imaging: a review. *Journal of Biomedical Optics*, 17(6):0612021–06120222, 2012.

[24] A. J. Devaney. A Computer Simulation Study of Diffraction Tomography. *IEEE Transactions on Biomedical Engineering*, BME-30(7):377–386, July 1983.

[25] Tian Ding, Kui Ren, and Sarah Valllian. A one-step reconstruction algorithm for quantitative photoacoustic imaging. *Inverse Problems*, 31(9):095005, September 2015.

[26] Belma Dogdas, David Stout, Arion F. Chatziioannou, and Richard M. Leahy. Digimouse: a 3d whole body mouse atlas from CT and cryosection data. *Physics in Medicine and Biology*, 52(3):577, February 2007.

[27] Neb Duric, Peter Littrup, Olivier Roy, Cuiping Li, Steve Schmidt, Xiaoyang Cheng, and Roman Janer. Clinical breast imaging with ultrasound tomography: A description of the SoftVue system. *The Journal of the Acoustical Society of America*, 135(4):2155–2155, April 2014.

[28] Neb Duric, Peter Littrup, Olivier Roy, Steven Schmidt, Cuiping Li, Lisa Bey-Knight, and Xiaoyang Chen. Breast imaging with ultrasound tomography: Initial results with SoftVue. In *IEEE International Ultrasonics Symposium*, pages 382–385. IEEE, July 2013.

[29] Nebojsa Duric, Peter Littrup, Lou Poulo, Alex Babkin, Roman Pevzner, Earle Holsapple, Olsi Rama, and Carri Glide. Detection of breast cancer with ultrasound tomography: First results with the Computed Ultrasound Risk Evaluation (CURE) prototype. *Medical Physics*, 34(2):773, 2007.

[30] Sergey A. Ermilov, Tuenchit Khamapirad, Andre Conjusteau, Morton H. Leonard, Ron Lacewell, Ketan Mehta, Tom Miller, and Alexander A. Oraevsky. Laser optoacoustic imaging system for detection of breast cancer. *Journal of Biomedical Optics*, 14(2):024007, 2009.

[31] Sergey A. Ermilov, Richard Su, Andre Conjusteau, Vassili Ivanov, Vyacheslav Nadvoretskiy, Tanmayi Oruganti, Pratik Talole, Fatima Anis, Mark A. Anastasio, and Alexander A. Oraevsky. 3d laser optoacoustic ultrasonic imaging system for research in mice (LOUIS-3dm). In *Proc. SPIE*, volume 8943, page 89430J, 2014.

[32] Todd N. Erpelding, Konstantin Maslov, Catherine Appleton, Julie A. Margenthaler, Michael D. Pashley, Jun Zou, Joseph P. Culver, Walter J. Akers, Samuel Achilefu, Dipanjan Pan, Gregory M. Lanza, and Lihong V. Wang. Superlinear clinical translation of photoacoustic tomography. In Robert J. Nordstrom, editor, *Translation research in biophotonics: four national cancer institute case studies*. SPIE, 2014.

[33] Netwon B. Everett, Barbara Simmons, and Earl P. Lasher. Distribution of blood (Fe50) and plasma (I131) volumes of rats determined by liquid nitrogen freezing. *Circulation Research*, 4:419–424, 1956.

[34] Paul L. Fackler. Notes on matrix calculus. *North Carolina State University*, 2005.

[35] Bengt Fornberg. The pseudospectral method: Comparisons with finite differences for the elastic wave equation. *Geophysics*, 52(4):483–501, 1987.

[36] Bengt Fornberg. *A Practical Guide to Pseudospectral Methods*. Cambridge University Press, October 1998.

[37] M. Galassi, J. Theiler, J. Davies, B. Gough, R. Priedhorsjy, G. Jungman, M. Booth, J.O. Tahtinen, T. Walter, F. Rossi, C. Perassi, S. Jaroszewicz, N. Darnis, J.H. Stover, I. Alxneit, T. Keskitalo, L. Johnson, P. Alken, R. Ulerich, P. Holoborodko, and P. Gonnet. GNU Scientific Library Reference Manual.

[38] Hartmut Gemmeke, Lea Althaus, Koen WA van Dongen, Herbert Egger, Jrgen Hesser, Jana Mayer, Nicole V. Ruiter, Michael Zapf, and Torsten Hopp. Wave equation based transmission tomography. In *2016 IEEE International Ultrasonics Symposium*, pages 1–4, 2016.

[39] J.F. Greenleaf, S.A. Johnson, R. C. Bahn, and B. Rajagopalan. Quantitative cross-sectional imaging of ultrasound parameters. In *Ultrasonics Symp Proc*, pages 989–995, 1977.

[40] Michelle Heijblom, Daniele Piras, Frank M. van den Engh, Margreet van der Schaaf, Joost M. Klaase, Wiendelt Steenbergen, and Srirang Manohar. The state of the art in breast imaging using the Twente Photoacoustic Mammoscope: results from 31 measurements on malignancies. *European Radiology*, 26(11):3874–3887, November 2016.

[41] Andrew J. Hesford and Weng C. Chew. Fast inverse scattering solutions using the distorted Born iterative method and the multilevel fast multipole algorithm. *The Journal of the Acoustical Society of America*, 128(2):679, 2010.

[42] T. Hopp, M. Zapf, E. Kretzek, J. Henrich, A. Tukalo, H. Gemmeke, C. Kaiser, J. Knaudt, and N. V. Ruiter. 3d ultrasound computer tomography: update from a clinical study. In *Proc. SPIE*, volume 9790, page 97900A, 2016.

[43] Ali Hormati, Ivana Jovanovic, Olivier Roy, and Martin Vetterli. Robust ultrasound travel-time tomography using the bent ray model. In *Proc. SPIE*, volume 7629, page 76290I, 2010.

[44] Yulia Hristova, Peter Kuchment, and Linh Nguyen. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Problems*, 24(5):055006, October 2008.

[45] Chao Huang, Liming Nie, Robert W. Schoonover, Zijian Guo, Carsten O. Schirra, Mark A. Anastasio, and Lihong V. Wang. Aberration correction for transcranial photoacoustic tomography of primates employing adjunct image data. *Journal of Biomedical Optics*, 17(6):066016, June 2012.

[46] Chao Huang, Kun Wang, Liming Nie, L.V. Wang, and M.A Anastasio. Full-Wave Iterative Image Reconstruction in Photoacoustic Tomography With Acoustically Inhomogeneous Media. *IEEE Transactions on Medical Imaging*, 32(6):1097–1110, June 2013.

[47] Chao Huang, Kun Wang, Robert W. Schoonover, Lihong V. Wang, and Mark A. Anastasio. Joint Reconstruction of Absorbed Optical Energy Density and Sound Speed Distributions in Photoacoustic Computed Tomography: A Numerical Investigation. *IEEE Transactions on Computational Imaging*, 2(2):136–149, June 2016.

[48] P. Huthwaite and F. Simonetti. High-resolution imaging without iteration: A fast and robust method for breast ultrasound tomography. *The Journal of the Acoustical Society of America*, 130(3):1721–1734, 2011.

[49] Steven L Jacques. Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, 58(11):R37–R61, June 2013.

[50] Finn B. Jensen, William A. Kuperman, Michael B. Porter, and Henrik Schmidt. Wave Propagation Theory. In *Computational Ocean Acoustics*. Springer-Verlag, New York, NY, 2011.

[51] Jithin Jose, Rene G. H. Willemink, Wiendelt Steenbergen, C. H. Slump, Ton G. van Leeuwen, and Srirang Manohar. Speed-of-sound compensated photoacoustic tomography for accurate imaging. *Medical Physics*, 39(12):7262–7271, December 2012.

[52] A. C. Kak. Computerized tomography with X-ray, emission, and ultrasound sources. *Proceedings of the IEEE*, 67(9):1245–1272, September 1979.

[53] Nathan Kaliss and David Pressman. Plasma and Blood Volumes of Mouse Organs, As Determined with Radioactive Iodoproteins. *Experimental Biology and Medicine*, 75(1):16–20, October 1950.

[54] Gordon S Kino. *Acoustic waves: devices, imaging, and analog signal processing*. Prentice-Hall, Englewood Cliffs, N.J., 1987.

[55] Andreas Kirsch and Otmar Scherzer. Simultaneous Reconstructions of Absorption Density and Wave Speed with Photoacoustic Measurements. *SIAM Journal on Applied Mathematics*, 72(5):1508–1523, January 2012.

[56] Thomas M. Kolb, Jacob Lichy, and Jeffrey H. Newhouse. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. *Radiology*, 225(1):165–175, October 2002.

[57] Jerome R. Krebs, John E. Anderson, David Hinkley, Ramesh Neelamani, Sunwoong Lee, Anatoly Baumstein, and Martin-Daniel Lacasse. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC177–WCC188, November 2009.

[58] Robert A. Kruger, Cherie M. Kuzmiak, Richard B. Lam, Daniel R. Reinecke, Stephen P. Del Rio, and Doreen Steed. Dedicated 3d photoacoustic breast imaging. *Medical Physics*, 40(11):113301, 2013.

[59] Geng Ku, Bruno D. Fornage, Xing Jin, Minghua Xu, Kelly K. Hunt, and Lihong V. Wang. Thermoacoustic and photoacoustic tomography of thick biological tissues toward breast imaging. *Technology in cancer research & treatment*, 4(5):559–565, 2005.

[60] R Lavarello and M Oelze. Density imaging using a multiple-frequency DBIM approach. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 57(11):2471–2479, November 2010.

[61] Roberto J. Lavarello and Andrew J. Hesford. Methods for Forward and Inverse Scattering in Ultrasound Tomography. In Jonathan Mamou and Michael L. Oelze, editors, *Quantitative Ultrasound in Soft Tissues*, pages 345–394. Springer Netherlands, Dordrecht, 2013.

[62] Roberto J. Lavarello and Michael L. Oelze. Density imaging using inverse scattering. *The Journal of the Acoustical Society of America*, 125(2):793, 2009.

[63] Yann Lecun. A Theoretical Framework for Back-Propagation. In *Proceedings of the 1988 Connectist Models Summer School*, pages 21–28, Pittsburgh, PA, 1988. CMU.

[64] Cuiping Li, Neb Duric, and Lianjie Huang. Clinical breast imaging using sound-speed reconstructions of ultrasound tomography data. In Stephen A. McAleavey and Jan D'hooge, editors, *Proc. SPIE*, volume 6920, page 692009, March 2008.

[65] Lei Li, Liren Zhu, Cheng Ma, Li Lin, Junjie Yao, Lidai Wang, Konstantin Maslov, Ruiying Zhang, Wanyi Chen, Junhui Shi, and Lihong V. Wang. Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution. *Nature Biomedical Engineering*, 1(5):0071, May 2017.

[66] Hongyu Liu and Gunther Uhlmann. Determining both sound speed and internal source in thermo- and photo-acoustic tomography. *Inverse Problems*, 31(10):105005, 2015.

[67] Yang Lou, Kenji Mitsuhashi, Catherine M. Appleton, Alexander Oraevsky, and Mark A. Anastasio. Generation of anatomically realistic numerical phantoms for optoacoustic breast imaging. In *Proc. SPIE*, page 97084O, March 2016.

[68] George D. Ludwig. The velocity of sound through tissues and the acoustic impedance of tissues. *J. Acoust. Soc. Am.*, 22(6):862–866, 1950.

[69] S. G. Mallat. *A wavelet tour of signal processing: the sparse way*. Elsevier/Academic Press, Amsterdam, 3rd edition, 2009.

[70] Srirang Manohar, Ren G. H. Willemink, Ferdi van der Heijden, Cornelis H. Slump, and Ton G. van Leeuwen. Concomitant speed-of-sound tomography in photoacoustic imaging. *Applied Physics Letters*, 91(13):131911, 2007.

[71] T. Douglas Mast, Laurent P. Souriau, D.-LD Liu, Makoto Tabei, Adrian I. Nachman, and Robert C. Waag. A k-space method for large-scale models of wave propagation in tissue. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 48(2):341–354, 2001.

[72] T. P. Matthews, K. Wang, C. Li, N. Duric, and M. A. Anastasio. Regularized Dual Averaging Image Reconstruction for Full-Wave Ultrasound Computed Tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64(5):811–825, May 2017.

[73] Thomas P. Matthews and Mark A Anastasio. Joint reconstruction of the initial pressure and speed of sound distributions from combined photoacoustic and ultrasound tomography measurements. *Inverse Problems*, 33:124002, 2017.

[74] K. Mitsuhashi, J. Poudel, T. Matthews, A. Garcia-Uribe, L. Wang, and M. Anastasio. A Forward-Adjoint Operator Pair Based on the Elastic Wave Equation for Use in Transcranial Photoacoustic Computed Tomography. *SIAM Journal on Imaging Sciences*, pages 2022–2048, January 2017.

[75] Jorge J. More and David J. Thuente. Line search algorithms with guatanteed sufficient decrease. *ACM Transactions on Mathematical Software*, 20:286–307, 1994.

[76] Philip M. Morse and K. Uno Ingard. *Theoretical Acoustics*. Princeton University Press, Princeton, NJ, 1986.

[77] Yurii Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259, August 2009.

[78] Jorge Nocedal. Updating quasi-Netwon matrices with limited storage. *Mathematics of Computation*, 35(151):773–782, 1980.

[79] Stephen J. Norton. Iterative inverse scattering algorithms: Methods of computing Frechet derivatives. *The Journal of the Acoustical Society of America*, 106(5):2653–2660, 1999.

[80] Stephen J. Norton and Melvin Linzer. Correcting for ray refraction in velocity and attenuation tomography: A perturbation approach. *Ultrasonic Imaging*, 4(3):201–233, July 1982.

[81] Brendan O'Donoghue and Emmanuel Candes. Adaptive Restart for Accelerated Gradient Schemes. *Foundations of Computational Mathematics*, 15(3):715–732, June 2015.

[82] Lauri Oksanen and Gunther Uhlmann. Photoacoustic and thermoacoustic tomography with an uncertain wave speed. *Math. Res. Lett.*, 21(5):1199–1214, 2014.

[83] Alexander A Oraevsky and Alexander A Karabutov. Optoacoustic Tomography. In Tuan Vo-Dinh, editor, *Biomedical Photonics Handbook*. CRC Press, Inc., Boca Raton, FL, 2003.

[84] Neal Parikh and Stephen Boyd. *Proximal algorithms*. Number 3 in Foundations and Trends in Optimization. Now Publishers, Inc., 2013.

[85] M. Perez-Liva, Joaquin L. Herraiz, J. M. Udias, Benjamin T. Cox, and Bradley E. Treeby. Full-wave attenuation reconstruction in the time domain for ultrasound computed tomography. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 710–713, 2016.

[86] Allan D. Pierce. Mathematical theory of wave propagation. In Malcolm J. Crocker, editor, *Encyclopedia of Acoustics*, volume 1. John Wiley & Sons, Inc., New York, NY, 1997.

[87] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, November 2006.

[88] Joemini Poudel, Thomas P. Matthews, Lei Li, Mark A. Anastasio, and Lihong V. Wang. Mitigation of artifacts due to isolated acoustic heterogeneities in photoacoustic computed tomography using a variable data truncation-based reconstruction method. *Journal of Biomedical Optics*, 22(4):041018, 2017.

[89] Joemini Poudel, Thomas P. Matthews, Kenji Mitsuhashi, Alejandro Garcia-Uribe, Lihong V. Wang, and Mark A. Anastasio. Iterative image reconstruction in elastic inhomogenous media with application to transcranial photoacoustic tomography. In Neb Duric and Brecht Heyde, editors, *Proc. SPIE*, page 101390C, March 2017.

[90] R. Gerhard Pratt, Lianjie Huang, Neb Duric, and Peter Littrup. Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data. In *Proc. SPIE*, volume 6510, page 65104S, March 2007.

[91] Jianliang Qian, Plamen Stefanov, Gunther Uhlmann, and Hongkai Zhao. An Efficient Neumann Series-Based Algorithm for Thermoacoustic and Photoacoustic Tomography with Variable Sound Speed. *SIAM Journal on Imaging Sciences*, 4(3):850–883, January 2011.

[92] Youli Quan and Lianjie Huang. Sound-speed tomography using first-arrival transmission ultrasound for a ring array. In *Proc. SPIE*, volume 6513, page 651306, 2007.

[93] N. V. Ruiter, M. Zapf, R. Dapp, T. Hopp, W. A. Kaiser, and H. Gemmeke. First results of a clinical study with 3d ultrasound computer tomography. In *2013 IEEE International Ultrasonics Symposium (IUS)*, pages 651–654, July 2013.

[94] G Y Sandhu, C Li, O Roy, S Schmidt, and N Duric. Frequency domain ultrasound waveform tomography: breast imaging using a ring transducer. *Physics in Medicine and Biology*, 60(14):5381–5398, July 2015.

[95] Nicol Schraudolph, Jin Yu, and Simon Gnter. A stochastic quasi-Newton method for online convex optimization. In *Proc. 11th Intl. Conf. Artificial Intelligence and Statistics*, volume 2, pages 436–443, 2007.

[96] J. S. Schreiman, J. J. Gisvold, J. F. Greenleaf, and R. C. Bahn. Ultrasound transmission computed tomography of the breast. *Radiology*, 150(2):523–530, February 1984.

[97] Laurent Sirgue and R. Gerhard Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, January 2004.

[98] Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors. *Optimization for machine learning*. The MIT Press, Cambridge, MA, 2012.

[99] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, New York, NY, USA, 2010.

[100] P. Stefanov and G. Uhlmann. Instability of the linearized problem in multiwave tomography of recovery both the source and the speed. *Inverse Problems & Imaging*, 7(4):1367–1377, 2013.

[101] Plamen Stefanov and Gunther Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011, 2009.

[102] Thomas L. Szabo. *Diagnostic Ultrasound Imaging: Inside Out*. Academic Press, December 2013.

[103] Makoto Tabei, T. Douglas Mast, and Robert C. Waag. A k-space method for coupled first-order acoustic propagation equations. *The Journal of the Acoustical Society of America*, 111(1):53–63, 2002.

[104] B.E. Treeby and B.T. Cox. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields. *Journal of Biomedical Optics*, 15(2):021314, 2010.

[105] Bradley E. Treeby, Trond K. Varslot, Edward Z. Zhang, Jan G. Laufer, and Paul C. Beard. Automatic sound speed selection in photoacoustic image reconstruction using an autofocus approach. *Journal of Biomedical Optics*, 16(9):090501, 2011.

[106] Bradley E Treeby, Edward Z Zhang, and B T Cox. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Problems*, 26(11):115003, November 2010.

[107] Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann. Seismic Waveform Inversion by Stochastic Optimization. *International Journal of Geophysics*, 2011:1–18, 2011.

[108] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, November 2009.

[109] Kun Wang, Thomas P. Matthews, Fatima Anis, Cuiping Li, Neb Duric, and Mark A. Anastasio. Waveform inversion with source encoding for breast sound speed reconstruction in ultrasound computed tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 62(3):475–494, 2015.

[110] Lihong V. Wang. Tutorial on Photoacoustic Microscopy and Computed Tomography. *IEEE Journal of Selected Topics in Quantum Electronics*, 14(1):171–179, 2008.

[111] Lihong V. Wang, editor. *Photoacoustic Imaging and Spectroscopy*. CRC Press, Inc., Boca Raton, FL, 2009.

[112] Lihong V. Wang and Hsin-i Wu. *Biomedical Optics: Principles and Imaging*. John Wiley & Sons, May 2007.

[113] J. Wiskin, D.T. Borup, S.A. Johnson, and M. Berggren. Non-linear inverse scattering: high resolution quantitative breast tissue tomography. *J. Acoust. Soc. Am.*, 131(5):3802–3813, 2012.

[114] Jun Xia, Muhammad R. Chatni, Konstantin Maslov, Zijian Guo, Kun Wang, Mark Anastasio, and Lihong V. Wang. Whole-body ring-shaped confocal photoacoustic computed tomography of small animals in vivo. *Journal of Biomedical Optics*, 17(5):050506, 2012.

[115] Jun Xia, Chao Huang, Konstantin Maslov, Mark A. Anastasio, and Lihong V. Wang. Enhancement of photoacoustic tomography by ultrasonic computed tomography based on optical excitation of elements of a full-ring transducer array. *Optics Letters*, 38(16):3140–3143, 2013.

[116] Lin Xiao. Dual averaging method for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596, 2010.

[117] Zhixing Xie, Fong Ming Hooi, J. Brian Fowlkes, Renee W. Pinsky, Xueding Wang, and Paul L. Carson. Combined Photoacoustic and Acoustic Imaging of Human Breast Specimens in the Mammographic Geometry. *Ultrasound in Medicine & Biology*, 39(11):2176–2184, November 2013.

[118] Zhixing Xie, Xueding Wang, Richard F. Morris, Frederic R. Padilla, Gerald L. Lecarpentier, and Paul L. Carson. Photoacoustic imaging for deep targets in the breast using a multichannel 2d array transducer. In Alexander A. Oraevsky and Lihong V. Wang, editors, *Proc. SPIE*, volume 7899, page 789907, February 2011.

[119] Qiaofeng Xu, Deshan Yang, Jun Tan, Alex Sawatzky, and Mark A Anastasio. Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam CT image reconstruction. *Medical Physics*, 43(4):1849–1872, 2016.

[120] Yuan Xu and Lihong V. Wang. Effects of acoustic heterogeneity in breast thermoacoustic tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 50(9):1134–1146, 2003.

[121] Yuan Xu and Lihong V. Wang. Time Reversal and Its Application to Tomography with Diffracting Sources. *Physical Review Letters*, 92(3), January 2004.

[122] Zhen Yuan, Qizhi Zhang, and Huabei Jiang. Simultaneous reconstruction of acoustic and optical properties of heterogeneous media by quantitative photoacoustic tomography. *Optics Express*, 14(15):6749–6754, 2006.

[123] Jin Zhang, Kun Wang, Yongyi Yang, and Mark A. Anastasio. Simultaneous reconstruction of speed-of-sound and optical absorption properties in photoacoustic tomography via a time-domain iterative algorithm. In *Proc. SPIE*, volume 6856, page 68561F, 2008.

[124] Zhigang Zhang, Lianjie Huang, and Youzuo Lin. Efficient implementation of ultrasound waveform tomography using source encoding. In *Proc. SPIE*, volume 8320, page 832003, February 2012.